Glutathione Transferases:

Probing for isoform specificity using Dynamic Combinatorial Chemistry

A thesis presented for the degree of

PhD in Biological Chemistry

by

Anne M. Caniard

University of Edinburgh

2010

Declaration

I herewith declare that all of the work and experiments in this thesis are my own unless otherwise noted. None of this work has been submitted in any other application for a higher degree.

Anne M. Caniard

Table of Contents

Abbreviations

ACE	Acetylcholine esterase
ATP	Adenosine triphosphate
bp	Base pair
BSA	Bovine serum albumin
CA	Carbonic anhydrase
CaM	Calmodulin
CAP	Catabolite activator protein
CDNB	1-chloro-2,4-dinitrobenzene
CMB	Chlorambucil
DCC	Dynamic combinatorial chemistry
DCL	Dynamic combinatorial library
DCM	Dichloromethane
DHLA	Dihydrolipoic acid
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EA	Ethacrynic acid
EDTA	Ethylenediaminetetraacetic acid
eEF	Eukaryotic elongation factor
EPNP	1,2-epoxy-3-(p-nitrophenoxy)propane
ESI-LC-MS	Electrospray-ionisation liquid-chromatography mass spectrometry
FRET	Fluorescence resonance energy eransfer spectroscopy
GDP	Guanosine diphosphate
GSH	Glutathione
G-site	Glutathione binding site
GSSG	Oxidised glutathione

GST	Glutathione transferase
GTP	Guanosine triphosphate
His_6	Hexahistidine
HOPDA	2-hydroxy-6-oxo-6-phenyl-2,4-dienoate
HPGDS	Hematopoietic prostaglandin D-synthases
HPLC	High pressure liquid chromatography
hr	Hour
H-site	Hydrophobic binding site
IC_{50}	Concentration of inhibitor generating 50% of inhibition
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal calorimetry
K	Dissociation constant
Kav	Gel phase distribution coefficient
k_{cat}	Turnover number
K_d	Enzyme-ligand binding constant
K_i	Inhibition constant
*K*M	Enzyme-substrate binding constant
LB	Luria Bertani
LC-MS	Liquid-chromatography mass spectrometry
MAAI	Maleylacetoacetate isomerase
MALDI-TOF	Matrix-assisted laser desorption/ionisation time-of-flight
MAP	Mercapturic acid pathway
MAPEG	Membrane-asssociated enzymes involved in eicosanoid and glutathione metabolism
min	Minute
Msu	1-menaphthyl sulphate
MWCO	Molecular weight cut-off
N	Stoichiometry of binding
NMR	Nuclear magnetic resonance

NTA	Nitrilotriacetic acid
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PGD	Prostaglandin D
PZQ	Praziquantel
QD	Quantum dots
R	Universal gas constant
RMSD	Root mean square deviation
RNA	Ribonucleic acid
sec	Second
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel elecrophoresis
T	Absolute temperature
TCHQ	Tetrachlorohydroquinone
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
tRNA	Transfer ribonucleic acid
UV	Ultra-violet
ΔH	Enthalpy
ΔS	Entropy

Acknowledgements

This research project would not have been possible without the help and support of several individuals who, in one way or another, contributed to the preparation and completion of this project.

First of all, I would like to thank Dr Dominic Campopiano and Dr Michael Greaney who have been my supervisors since the beginning of this study. Thanks for guiding me in my work, and for letting me the freedom to take initiatives in the development of my lines of research.

I gratefully acknowledge Prof. Alan Cooper and Prof. Mark Bradley for their advice and constructive comments throughout this project.

Thanks to Dr Dave Clarke, Dr Marika De Cremoux, and Margaret Nutley for their guidance in different aspects of my research and for their technical assistance.

This project would not have been possible without the collaboration of Venu Bhat who worked with me on this project from the start. He was always full of enthousiasm and resourceful ideas. It was a great pleasure working with him.

I warmly acknowledge the group members of Lab 229. Special thanks to Karin Bodewits and Philipp Grämlich for their presence, their constructive comments on this thesis, and for providing me a home when I needed it. Thanks to Géraldine Eicher for her precious help.

Finally, I want to express my gratitude to my family who always encouraged me to concentrate on my study. Thanks to Swan for his continual moral support through the preparation of this thesis.

A special thought to my Grand-father to whom I dedicate this thesis.

Abstract

Cytosolic glutathione transferases (GSTs) are a large family of enzymes that play an important role in detoxification of xenobiotics. They catalyse the conjugation of the glutathione tripeptide (GSH) to a wide range of toxic electrophilic acceptors. The overall 3D folds and architectures of the catalytic sites of many GSTs are conserved. They are composed of a well conserved glutathione binding site (G-site) and a promiscuous hydrophobic binding site (H-site). The 3D structure and ligand specificity has allowed the sub-classification of the multiple isoforms within the soluble GST superfamily. GSTs are involved in the drug detoxification and so are the target of medicinal chemistry programmes but it has proven difficult to generate isoform-specific inhibitors due to their inherent promiscuity.

In this project, Venughopal Bhat (University of Edinburgh, laboratory of Dr. Mike Greaney) and I have explored a new platform to probe enzyme specificity. Protein-directed dynamic combinatorial chemistry (DCC) allows the assembly and amplification of a ligand within the confines of a binding site. DCC was used as a tool to explore the promiscuous H-site of four eukaryotic GSTs. I purified recombinant forms of SjGST, hGST P1-1, mGST M1-1 and mGST A4-4 from *E. coli* and assayed them with the universal, synthetic GST substrate 1-chloro-2,4-dinitrobenzene (CDNB). Venughopal Bhat prepared a ten-member, thermodynamically-controlled, dynamic combinatorial library (DCL) of acyl hydrazones from a 1-chloro-2-nitrobenzene aldehyde and ten acylhydrazides. This DCL was incubated with each of the four GST isozymes (spanning diverse classes) and distinct amplification effects were observed for SjGST and hGST P1-1. I subsequently carried out several biophysical experiments in an attempt to rank each of the ligands. These experiements, coupled with molecular modelling, provided insight into the basis of the observed selectivity.

Bacterial GSTs are thought to play a role in primary metabolism and display a different GSH-conjugation mechanism compared to the eukaryotic GSTs. A recombinant form of the beta-class GST from the pathogenic bacterium *Burkholderia*

cenocepacia was isolated, purified and biochemically characterised. The same ten-member acylhydrazone DCL was interfaced with the bacterial GST which was shown to amplify a hydrophobic library member that shared structural features with the known substrate 2-hydroxy-6-oxo-6-phenyl-2,4-dienoate (HOPDA).

With the collaboration of Venughopal Bhat, I attempted to explore the putative active site of a GST-like protein with an unknown function using the same DCL. Although no amplification was observed, a new aldehyde template was suggested for future DCC experiments on this protein.

GSTs are widely employed in biotechnology as protein fusion tags to enhance target protein solubility coupled with a facile enzyme assay. Manish Gupta and Juan Mareque-Rivas (University of Edinburgh) used the N-terminal, hexahistidine-tagged SjGST to demonstrate that quantum dots (QDs) coated with nitrilotriacetic acid (NTA) bound to Ni^{2+} ions can be used to reversibly and selectively bind, purify, and fluorescently label a His_6-tagged GST in one step with retention of enzymatic activity. For this prupose, I purified and characterized both the untagged and hexahistidine-tagged – SjGST prior to their experiments.

Chapter 1: Probing the active site of four eukaryotic glutathione transferases using dynamic combinatorial chemistry

1.1. Introduction

1.1.1. Introduction to glutathione transferases

1.1.1.1. Glutathione transferases – role in cellular detoxification

All living organisms are open systems constantly interacting with their environment. They can take up various nutrient compounds required for growth which are metabolised through catabolic pathways which degrade these molecules to liberate their energy and produce small building blocks. Anabolic pathways integrate them into newly synthesized macromolecules. All these biochemical reactions are rigorously controlled by various enzymes. Waste and heat are released back to the environment.

A xenobiotic is a chemical which enters a cell but has no physiological function, or an endogenously produced substance (endobiotic) present in a higher concentration than expected. Drugs, some food constituents, cosmetics, toxins, pollutants are considered as xenobiotics in humans because they are not synthesized by the human body, nor are they essential nutrients. The body removes xenobiotics by detoxication and secretion of the resulting substances through urine, feces, sweat or breath [1]. The metabolism of xenobiotics can be divided into three phases: (I) activation, (II) conjugation and (III) excretion. Phase I and phase II involve the conversion of a hydrophobic xenobiotic molecule to a relatively more water-soluble and therefore less toxic metabolite, which can then be easily eliminated from the cell (phase III). In phase I, a variety of enzymes, mainly cytochrome P450 systems, creates or modifies a functional group on their substrates. These reactions include reduction, oxidation, hydrolysis and hydration [2]. In phase II, which can also occur without or before phase I, xenobiotics are conjugated to endogenous more water-soluble substrates such as glutathione (GSH), glucuronic acid or glycine. These reactions are catalysed by several groups of transferases [1, 3-7]. Glutathione Transferases (also known as Glutathione S-Transferases; GSTs) constitute the most important group of transferases, and catalyse the conjugation of glutathione to a wide range of electrophilic substrates. Once conjugated, xenobiotics can be further metabolized: in particular, glutathione

conjugates are processed to acetylcysteine conjugates through the mercapturic acid pathway (MAP) [8]. In phase III, transformed xenobiotics are released from cells through a variety of membrane ATP-dependant pump systems [9].

1.1.1.2. The mercapturic acid pathway

The body can detoxify many xenobiotics via the glutathione-dependant mercapturic acid pathway (Figure 1.1.), through inter-organ cooperation between the liver, kidney and intestine. This starts with the formation of a thioether of glutathione, which is ultimately converted to a mercapturic acid (S-substituted N-acetylated-L-cysteine). Mercapturic acids are organic acids, and are therefore more water-soluble than the entering xenobiotics. Their higher solubility facilitates their excretion [8].

Firstly, whereby an electrophilic molecule is attacked by the tripeptide glutathione, is catalysed by GSTs. These enzymes are found in all tissues and cell types but are more abundant in the liver. They bring together the electrophilic centre of a xenobiotic with the activated thiolate of glutathione. Their glutathione binding site is well conserved throughout the family but they have a broad specificity towards the hydrophobic xenobiotics [8, 10-12]. The GST super-family will be presented and discussed further in the following sections.

The conjugated compound is exported from the cell by an ATP-dependent transport and is either excreted in the biliary canniculae or transported to the kidney. Here the glutathione moiety is degraded to a cysteine in a two-step mechanism. γ-glutamyl transpeptidase – an extra-cellular membrane-bound enzyme – catalyses the first step by removing a glutamate. The isopeptide bond between the γ-carboxyl of glutamate and the amino group of cysteine is cleaved, and the γ-glutamyl is either hydrolysed or transferred to an acceptor amino-acid and a cysteinylglycine conjugate remains [8, 13-16].

The second step in the degradation of the glutathione moiety is also thought to happen extracellularly in the biliary canniculae and kidney and is catalysed by a

membrane-associated cysteinylglycine dipeptidase. The glycyl is hydrolised, leaving an *S*-substituted cysteine conjugate [8, 17].

The remaining conjugate is *N*-acetylated to give an *N*-acetylated-*S*-cysteine conjugate (mercapturic acid). This reaction occurs intracellularly mainly in the kidney and is catalysed by a microsomal cysteine *S*-conjugate *N*-acetyltransferase that transfers an acetyl group from acetylcoenzyme A to the amino group of the cysteine conjugate. Finally, the mercapturic acid is transported out of the cell and excreted in urine [8, 18].

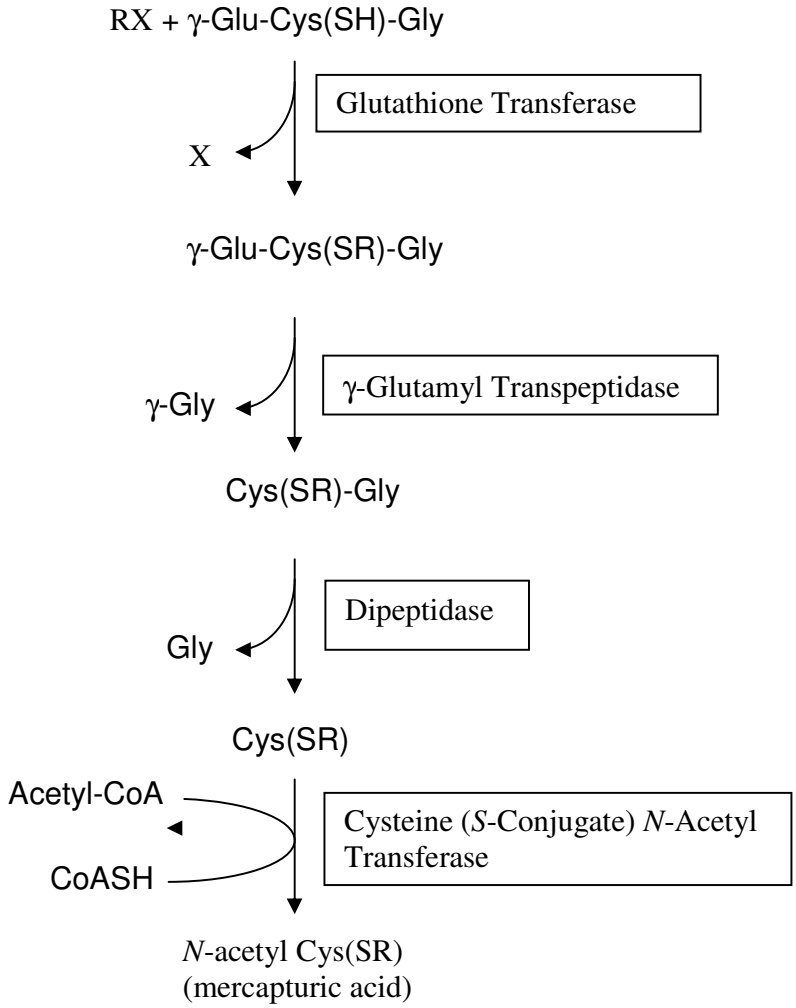


Figure 1.1. Sequence of enzymatic steps in mercapturic acid pathway

1.1.1.3. Glutathione: an unusual peptide

1.1.1.3.1. The structure of glutathione

Glutathione – γ-L-glutamyl-L-cysteinylglycine, GSH (Figure 1.2.) – is a natural antioxidant which plays a crucial role in detoxification processes; in particular, it largely contributes to the mercapturic acid pathway described above.

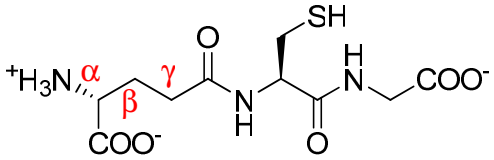


Figure 1.2. The structure of L-glutathione

This tripeptide was first isolated in 1921 by Hopkins [19] and its structure was determined by Nicolet in 1930 [20]. It is widely distributed throughout animal tissues and plants at high cellular concentrations (~1-10 mM) and is especially concentrated in the liver (~5-10 mM), which makes GSH the principal cellular non-protein thiol and the most abundant low molecular weight peptide in eukaryotes. In bacteria, GSH is mostly present in Gram-negative aerobic species (~3.5-6.6 mM in *E. coli*) but very rarely in anaerobes and Gram-positive strains which have developed other low molecular weight thiols such as mycothiol produced by *Actinobacteria* or bacillithiol found in *Bacilli* [21-23].

GSH presents some unique properties which are closely related to its functions:

- Its characteristic structural features, including the unusual peptide linkage between the amine group of L–cysteine and the carboxyl group of the L–glutamate side chain, provides for intra-cellular stability and specificity in glutathione-dependent enzymes, particularly with GSTs;

- Its cysteinyl residue (pK_a of GSH = 9.2) provides a nucleophilic thiol which allows the formation of conjugates with both endogenous and exogenous compounds;

- Its overall negative charge and hydrophilicity increase the solubility of the xenobiotics to which it becomes conjugated [12, 24-27].

1.1.1.3.2. Glutathione metabolism

GSH metabolism occurs through the γ-glutamyl cycle which can be divided in six steps.

GSH synthesis takes place in the cytoplasm and involves two ATP-dependant enzymatic steps (Figure 1.3.). The first step is rate-limiting and consists in the condensation of an L-glutamate and an L-cysteine. This step is limited by the feedback of competitive inhibition by glutathione and by the availability of L-cysteine. This reaction is catalysed by a glutamate cysteine ligase (GCL) which requires either Mg^{2+} or Mn^{2+} and coupled ATP hydrolysis. L-glutamate and ATP first react to give a tightly bound γ-glutamyl phosphate which reacts with L-cysteine. Therefore an amide bond is formed between the γ-carboxyl group of glutamate and the amino group of cysteine. GCL is a dimeric enzyme composed of a heavy unit with competent but low catalytic activity and a modulatory activity light subunit lowering the K_m for glutamate and increasing the K_i for glutathione. GSH cell levels are closely related to GCL expression and activity. Oxidative stress caused by diverse agents can transcriptionally and post-transcriptionally activate the GCL subunits. Increases in GSH levels, GCL mRNA levels and GCL activity are also observed in drug-resistant tumor cell lines [24, 26, 28-32].

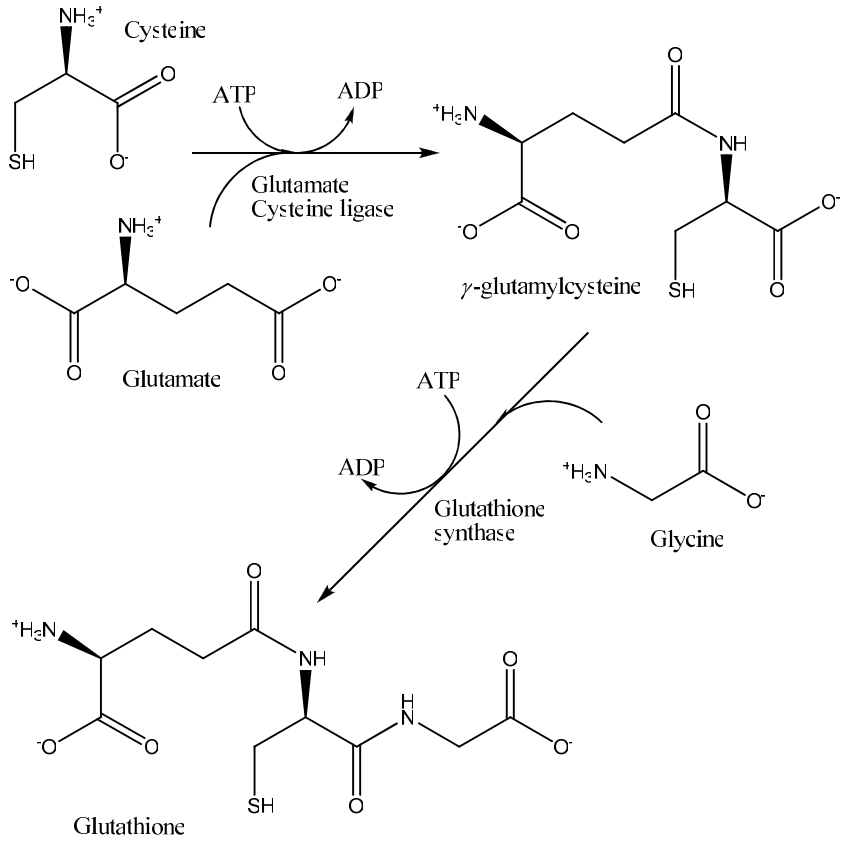


Figure 1.3. Glutathione biosynthesis mechanism

The second step of glutathione biosynthesis is catalysed by a glutathione synthase for which chemical mechanisms are similar to those of GCL: L-g-glutamyl-L-cysteine and ATP first react to form tightly bound L-g-glutamyl-L-cysteinylphosphate, and that intermediate then reacts with glycine to form GSH [24, 26, 28-30, 33].

The breakdown of GSH begins outside of the cell and is catalysed by the γ-glutamyl transpeptidase involved in the MAP. The γ-glutamyl moiety is transferred to acceptors such as amino acids, certain dipeptides and water. γ-glutamyl amino acids are then transported into the cell where they become substrates of γ-glutamyl cyclotransferase, which converts these products into 5-oxo-L-proline. The ATP dependant conversion of 5-oxo-L-proline to L-glutamate is catalysed by the intracellular enzyme 5-oxo-prolinase.

The cysteinylglycine formed in the transpeptidase reaction is split by dipeptidase to generate cysteine and glycine [24, 28]. Most cells readily take up cysteine and the majority is incorporated into GSH, while the rest is either incorporated into proteins or degraded into sulphate and taurine. The γ-glutamyl cycle allows GSH to be used as a continuous cellular source of cysteine [26].

1.1.1.3.3. Glutathione functions

In eukaryotic cells, GSH is mostly stored in the cytoplasm (almost 90%) where it serves several vital functions, 10% is found in mitochondria and a small percentage is in the endoplasmic reticulum [34, 35].

1.1.1.3.3.1. Antioxidant function

The main function of glutathione is to protect cells against reactive oxidants. All aerobic organisms are subject to oxidative stress from the respiration metabolism. Reactive oxygen species, including free radicals such as superoxide, hydroxyl radicals and non-radical species like hydrogen peroxide are formed and can lead to lipid peroxidation and cellular damage.

GSH primarily exists in two redox forms: reduced and oxidised glutathione disulfide (GSSG). The latter only represents less than one percent of the total glutathione, it can be reduced back to GSH at the expense of NADPH through the action of a glutathione reductase. When this reduction is overcome by a severe oxidative stress, GSSG can be actively exported out of the cell to prevent a shift in the redox equilibrium. The presence of a sulfhydryl group makes GSH function as an antioxidant; thiols react rapidly and non-enzymatically with oxidising species and some of their oxidation products can be reversibly reduced by various redox-active enzymes such as thioredoxin and glutaredoxin.

1.1.1.3.3.2. Detoxifying function

As described previously in section 1.1.1.2., glutathione plays a major role in the detoxification of many exogenous and endogenous compounds through the mercapturic acid pathway. Conjugation to xenobiotics irreversibly consumes GSH [26].

1.1.1.3.3.3. Maintenance of the intracellular redox state and cell signalling

GSH is the most abundant intracellular low-molecular-mass thiol and therefore is an important mediator in the maintenance of the intracellular redox state and the essential thiol status of proteins. It may be thought of as an intracellular redox buffer which can act as an electrophile scavenger as well as an electron donor via the sulfhydryl group of a cysteine residue.

Glutathione undergoes protein disulfide exchange as follows:

$$\text{Protein-SSG} + \text{GSH} \quad \leftrightarrow \quad \text{Protein-SH} + \text{GSSG}$$

This reversible reaction is catalysed by a thiol transferase and therefore the equilibrium depends on the concentrations of GSH and GSSG. During oxidative stress, disulfide exchange between proteins and glutathione constitutes a protective mechanism for thiols, preventing further oxidation, and plays roles in cell signaling. In particular, protein S-glutathionylation is known to regulate several metabolic processes including redox homeostasis, ion channel activity, protein folding, cell growth and apoptosis [21, 26, 30, 36-42].

1.1.1.4. The glutathione transferase super-family

Glutathione Transferases (EC 2.5.1.18, GSTs) are a large family of phase II detoxification enzymes which catalyses the first step in the mercapturic acid pathway.

The general chemical mechanism of these enzymes is thought to involve nucleophilic attack by the thiol of glutathione on an electrophilic acceptor substrate (Figure 1.4.) [43].

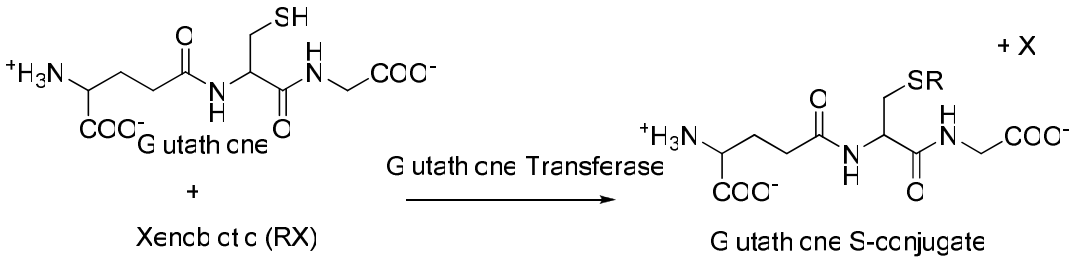


Figure 1.4. Glutathione conjugation to a generic xenobiotic substrate (X) catalysed by GST results in the formation of a glutathione-S-X conjugate.

Typical GST-catalyzed reactions are varied, and include Michael-type addition with ethacrynic acid (EA), nucleophilic addition to the epoxide phenantrene 9,10-oxide, hydroperoxide reduction with cumene hydroperoxide, and thiolysis of 4-nitrophenyl acetate to name a few (Figure 1.5.) [44].

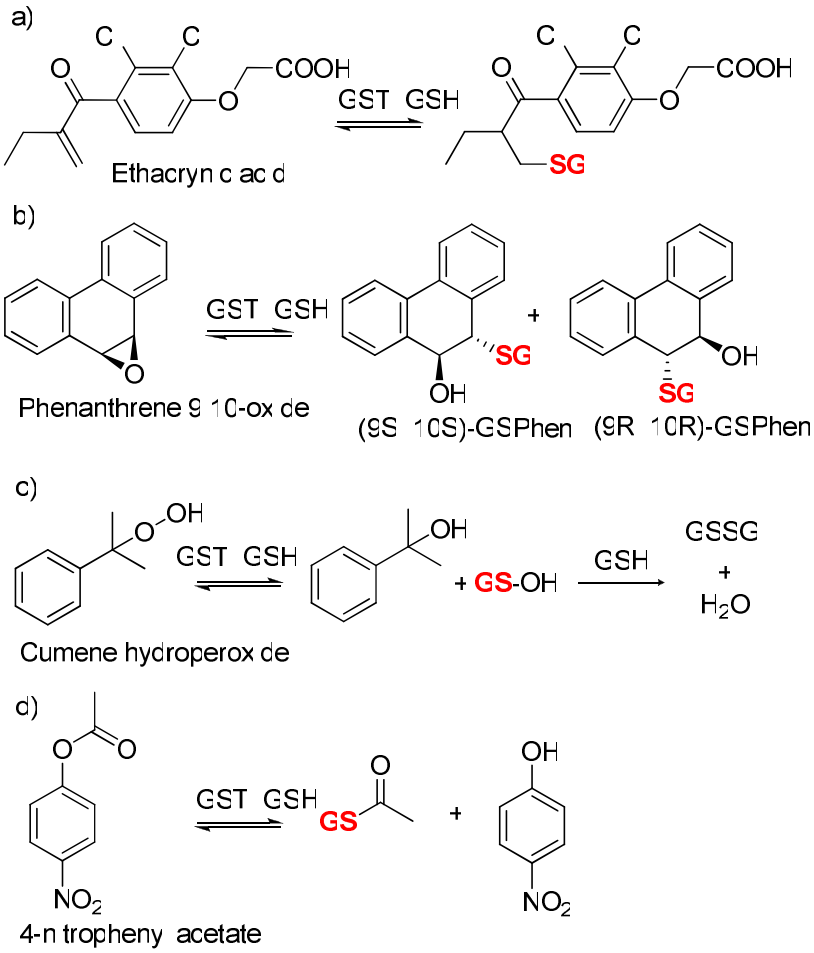


Figure 1.5. Examples of reactions catalysed by GSTs: a) Michael-type addition; b) Nucleophilic addition; c) Hydroperoxide reduction d) Thiolysis.

GST assay

Most GSTs can catalyse the conjugation of GSH to 1-chloro-2,4-dinitrobenzene (CDNB; Figure 1.6.), which is considered as the 'universal GST substrate'. It has been used since 1974 to monitor the enzymatic activity of various GST by visible spectroscopy, as the conjugate resulting from the reaction absorbs at 340 nm (ε = 9600 $M^{-1}.cm^{-1}$) [43].

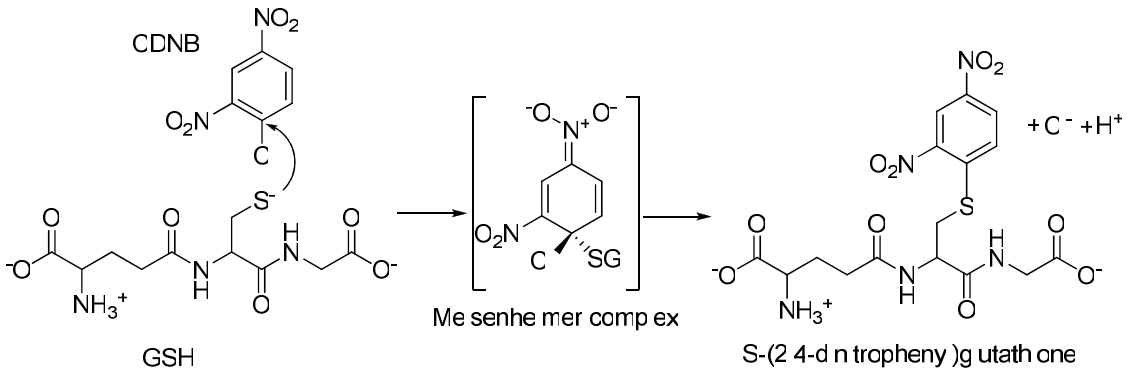


Figure 1.6. Conjugation of glutathione and CDNB catalysed by many GSTs.

GSTs were discovered in 1961 by Booth *et al.*, who partially purified and characterised a cytosolic GST isolated from rat liver [45]. Since then, GSTs have been extensively studied and classified according to a variety of criteria including protein sequences, kinetic and ternary/quaternary structural properties. There are four families of GSTs: two of those, cytosolic and mitochondrial, are soluble and distantly related. The third family, microsomal, is referred to membrane-associated enzymes involved in eicosanoid and glutathione metabolism (MAPEG). The fourth family is less related than the three other and is represented by the bacterial fosfomycin resistance proteins FosA and FosB [46-50]. Recent, detailed reviews have been published on the GST family so in this thesis I will highlight GSTs of interest to this study.

1.1.1.4.1. Cytosolic glutathione transferases

Cytosolic GSTs represent the largest family and are found in a wide range of species including mammals, plants, invertebrates, insects, fungi and bacteria. Historically, mainly based on amino-acid sequence similarity, they have been divided into seven classes: Alpha, Mu, Omega, Pi, Sigma, Theta, Zeta. The letter designation of the enzymes (A, M, O, P, S, T and Z, respectively) shows their assignments to one of these classes and their subunit composition or isoenzyme type is designated by Arabic

numerals. Single-letter prefixes denote species of origin (e.g. m, mouse; h, human). For instance, a homodimer of type 4 alpha subunits from the mouse is mGSTA4-4. Other classes: Beta, Delta, Epsilon, Lambda, Phi, Tau and the "U" class have been identified in plants, insects and bacteria. The plant and bacterial enzymes often serve more specialised functions than detoxification, for instance they can take part in the transport of pigments or dehalogenation of hydrocarbons [46-53].

1.1.1.4.1.1. Protein architecture of cytosolic glutathione transferases

In general, GST monomers are ~25-30 kDa and the amino acid sequence identity within a class is high (more than 40%) whereas the identity between classes is much lower (less than 30%). Cytosolic GSTs are mostly homodimers with a common three–dimensional structural fold. Each monomer contains two main domains connected by a short linker region (Figure 1.7.).

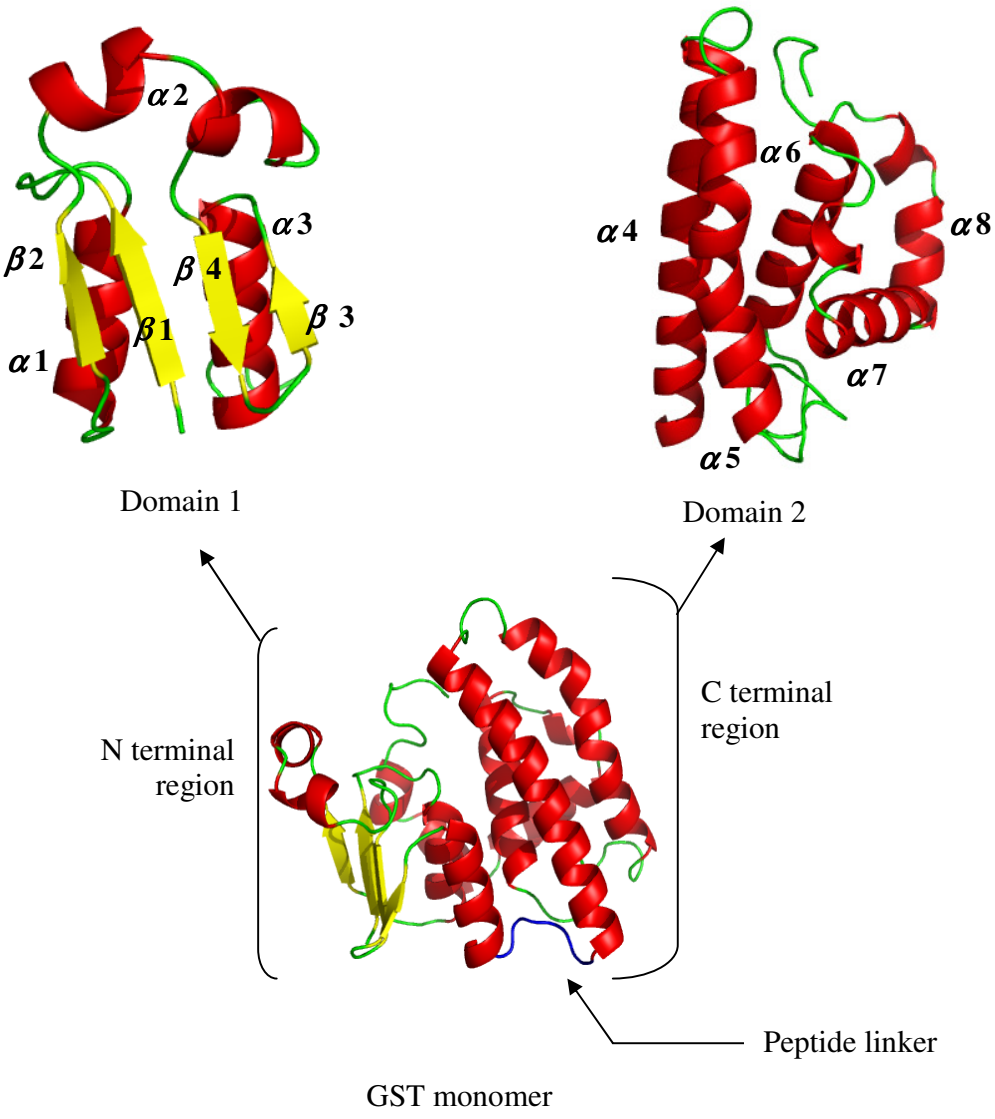


Figure 1.7. Crystal structure of a ligand-free hGSTP1 monomer (PDB 1EOG). α helices and β strands are represented in red and yellow, respectively. The peptide linker connecting the two domains is shown in blue.

The N-terminal domain or domain 1 constitutes roughly one-third of the protein and is classified as part of the thioredoxin superfamily fold [54]. It is composed of three α helices and a mixed β sheet of the sequence *β1-α1-β2-α2-β3-β4-α3*. These elements fold so that the β strands are in the order of 4-3-1-2 with strand 3 antiparallel to the

others. The β sheet is almost planar with α1 and α3 helices below this plane and α2 above it. There is a loop with a proline residue in the *cis* conformation between α2 and β3 which helps to maintain the correct structure for catalytic activity [55]. Most GSTs present a SNAIL/TRAIL motif (*i.e.* Ser-Asp-Ala-Ile-Leu/Thr-Arg-Ala-Ile-Leu) in the α3 helix, which includes residues contributing to the GSH binding site.

The larger C-terminal domain or domain 2 is an all-α-helical domain, the core of which generally consists of a group of five helices (the number of α helices varies widely between the classes). This domain is less conserved at the sequence and structural levels, with the helices varying in number, length, curvature, and orientation.

The dimeric structures of GSTs (~ 50 kDa) are very stable and class-specific. The subunits are non-covalently bound and are related by a two-fold axis (C_2 symmetry). Interactions between the two subunits are mainly hydrophobic but differ between the classes. GST dimerisation allows for the construction of two fully functional active sites which are partly located at the dimer interface [47, 49, 50, 56].

1.1.1.4.1.2. G- , H- and L- sites

GSTs are highly specific towards GSH and relatively more flexible with the electrophilic substrate. Each GST active site can be divided into a GSH binding site or "G-site" and a hydrophobic substrate binding site or "H-site" (Figure 1.8.).

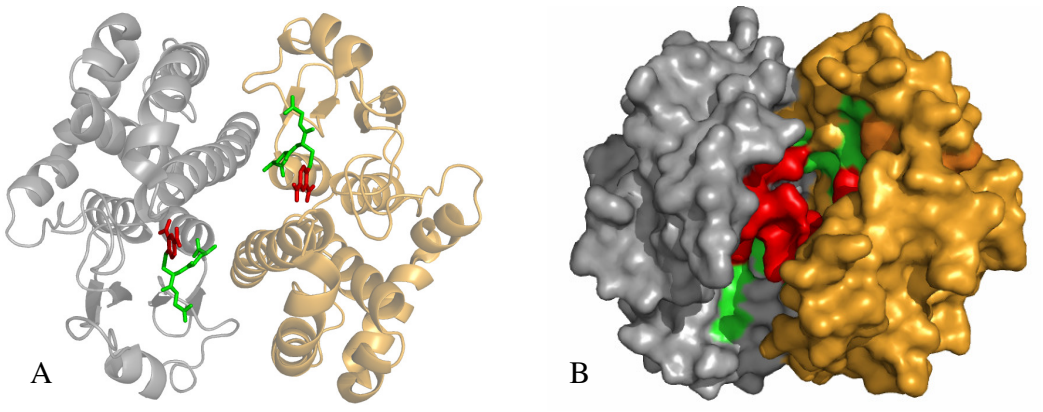


Figure 1.8. A. Structure of hGSTM1-1 with one molecule of glutathionyl-S-dinitrobenzene bound in each active site (grey: monomer 1; orange: monomer 2; green: glutathione moiety; red: dinitrobenzene moiety). Structure of the same GST illustrating the G- and the H- sites (grey: monomer 1; orange: monomer 2; green: G-site; red: H-site; PDB 1XWK) [57]. This Figure was generated using PyMOL.

G-site and activation of glutathione

The G-site is located in a cleft between the N-terminal domain of one subunit and the C-terminal domain of the other subunit, so it is only completed after dimerisation. Most of the G-site residues are located in domain 1 and involve a specific network of polar bonds with GSH (Figure 1.9.). Amino-acids contributing to binding of GSH at the G-site are generally well conserved within classes. The tripeptide binds GSTs in an extended conformation with its main chain forming hydrogen-bonds with the peptide-loop connecting $\alpha 2$ to $\beta 3$. This interaction is important for the recognition and orientation of GSH in the active site. The γ-glutamyl moiety anchors the tripeptide in the G-site by interacting extensively with a hydrophilic pocket located near the subunit interface. The glycyl resides near the solvent-accessible surface of the protein, with its carboxyl group strongly sequestered at the G-site. Finally, the cysteinyl sulfur points toward the subunit to which it is bound. Depending on the class of GST, the thiol of GSH is thought to be activated and stabilised by a tyrosine, a serine or a cysteine residue

(Figure 1.10.). Within the π and α classes, it was found that mutation of a tyrosine to a phenylalanine dramatically decreased the catalytic activity but did not significantly reduce the ability to bind GSH, suggesting that this interaction does not have an essential binding-role. [46, 56, 58, 59].

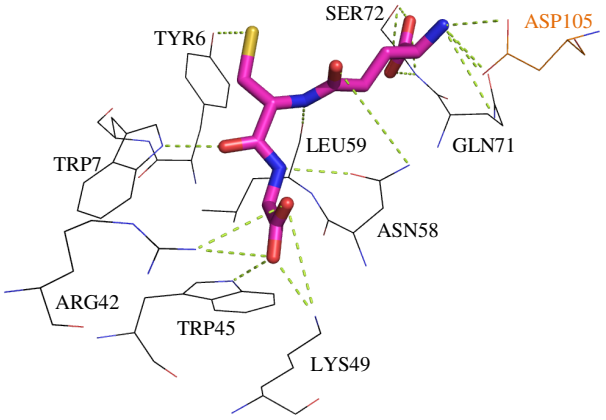


Figure 1.9. Representation of the hGSTM1-1 G-site residues interacting with glutathione (PDB 1XW6) through polar bonds. (black: amino acids from domain 1 of monomer 1; orange: Asp105 from domain 2 of monomer 2; violet: glutathione) [57]. This Figure was generated using PyMOL.

As GSH binds to the G-site, the pK_a value of the thiol drops from ~9.5 to ~6.8 units, which promotes deprotonation of the GSH at physiological pH (~7.4). Proton release and extrusion from the active site takes place via an initial conformational rearrangement of GSH, followed by a water-assisted proton transfer between the thiol group of GSH cysteinyl residue and the GSH glutamyl α-carboxyl group. When the catalytic residue is a tyrosine or a serine, the anionic form of GSH is stabilised by the available hydroxyl group, and can subsequently react more readily with electrophilic substrates [56, 60]. GSTs possessing this tyrosine have higher activity toward the GSH conjugation reaction than GSTs containing a serine [61] as the hydroxyl group of a tyrosine is more acidic (pKa ~10.5) than the hydroxyl group of a serine (pKa ~13). Therefore, the anionic form of GSH would make stronger interaction with the tyrosine, resulting in a higher stability of the activated substrate.

A theta-class GST (hGSTT2-2) with a catalytically active serine residue has been demonstrated to act as a sulfatase with menaphthyl sulphate, which generates menaphthyl-GSH and free sulphate [62]. Although the enzymes residues involved in this catalysis have not been clearly identified, it has been shown that the active site serine is not essential in this reaction. The architecture of the active site could contribute to catalysis in this isozyme. Therefore, these enzymes may have evolved to specifically catalyse certain types of reactions [53].

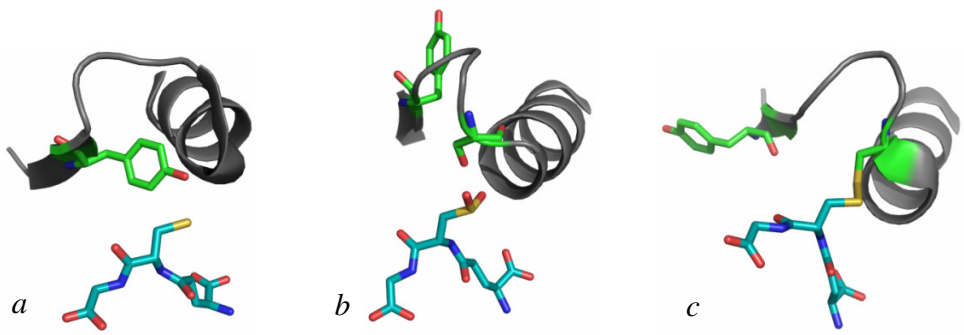


Figure 1.10. Catalytic residues in GSTs. In most GST classes, a tyrosine residue (*a*: hGSTM1 crystal structure complexed with glutathione; 1XW6) interacts with GSH to stabilize the thiolate anion, with a consequent decrease in p*K*a. In the theta, kappa and the zeta classes, this role is carried out by a serine residue (*b*: hGSTK1 crystal structure complexed with glutathione sulfinate; 1YZX), while in the omega and beta classes a mixed disulphide is formed with a cysteine residue (*c*: hGSTO1 crystal structure complexed with glutathione; 1EEM). This figure was generated using PyMOL.

The mechanism of GSH activation by a catalytic tyrosine is not well understood. Nishida and co-workers [63] highlighted the presence of an additional conserved histidine residue in the active sites of beta-class GSTs. They suggested that the sulphur atom of the GSH thiol group makes hydrogen bonds with the backbone NH and the SH group of the essential cysteine and with the imidazole ring of the nearby histidine (His106 in the *E. coli* GST; [63]). The deprotonation of the GSH thiol group would be accelerated by either the imidazole ring of the histidine or by the sulphur atom of the cysteine.

GSTs containing a cysteine in their active site often show dehalogenation activity. In general, this reaction is a two-step mechanism [64]:

(1) A molecule of GSH is added to a halogenated substrate, and the halogen atom is eliminated. The dehalogenated molecule is released by nucleophilic attack of the active site cysteine thiolate on the sulphur of the GSH moiety, yielding a mixed disulfide linkage;

(2) Regeneration of the enzyme is initiated by the binding of a second GSH at the H-site for a final disulfide exchange. The result is a reduced catalytically active cysteine thiol and the release of GSSG from the active site.

H-site and the binding of electrophiles

In contrast to GSH binding to domain 1, residues from domain 2 are mainly responsible for binding of the electrophilic substrates. The H-site is highly hydrophobic and includes the active-site loop connecting β1 to α1, the C-terminal region of α4 and the C-terminal segment of the peptide linker. The great diversity of H-site topologies would explain the differences in electrophilic substrate specificity between GST classes [46, 56].

L-site or ligandin binding site

In addition to their enzymatic properties, GSTs are able to bind hydrophobic molecules, including hemin, bilirubin, bile salts, steroids, thyroid hormones, fatty acids and drugs. This ligandin property was observed in early work and suggests a role of storage or transport [65, 66]. The ligandin binding site or "L-site" has been observed in many GST isoforms. Its location remains uncertain and seems to be variable among the classes. For instance, McTigue *et al.* have shown that the drug Praziquantel could bind to the dimer interface of a GST from the parasitic worm *Schistosoma japonicum*, whereas the human GST P1-1 L-site would be located at a buffer binding site in the H-site [67-69].

1.1.1.4.2. Glutathione transferases as therapeutic targets

1.1.1.4.2.1. Mammalian cytosolic glutathione transferases and cancer

Development of acquired drug resistance is a major problem in chemotherapy treatment. Exposure to anti-cancer agents can lead to an induction and expression of gene products that protect the cells. In particular, over-expression of GST isozymes has been reported in a number of different human cancers when compared to normal tissues, and it has been demonstrated that it is correlated with clinical drug resistance. GSTs can serve two distinct roles in the development of drug resistance: they can increase the rate of direct drug detoxification; and they can regulate the mitogen-activated protein kinase pathway that participates in cellular survival and apoptosis signalling [70-73].

1.1.1.4.2.1.1. Mammalian cytosolic glutathione transferases

Three-dimensional ribbon structural representations of the mammalian GSTs discussed below are shown in Figure 1.12. Isoform-specific features are summarised in Table 1.1.

1.1.1.4.2.1.1.1. Alpha-, mu- and pi-classes

The Alpha, Mu and Pi classes are the most abundant in mammals and the most extensively studied GSTs. Based on sequence alignments and substrate specificity criteria, several isoenzymes were recognized as belonging to Alpha, Mu or Pi class. In humans, four distinct subunits have been characterized in the Alpha class (A1 to A4), four in the Mu class (M1 to M4) and only one belongs to the Pi class (P1). Mammalian Alpha/Mu/Pi overall structures are very similar, they share 26 conserved residues (pair-wise sequence identity: alpha-mu 20%, pi-mu 30 % and alpha-pi 32 %), but they show some important points of difference in the detailed structure, in particular in domain 2 [46, 49, 53, 56].

The three classes share the same type of hydrophobic "lock and key" interaction between the subunits 1 and 2. The 'key' is an aromatic residue (Phe-52 in Alpha, Phe-56 in Mu and Tyr-49 in Pi) extending from the loop preceding β-3. It fits into a hydrophobic 'lock' composed by helices α-4 and α-5 of the other subunit (Figure 1.11.). The interface creates a V-shape crevice which is solvent-accessible. Mu-class enzymes have a characteristic mu-loop which is located between β-2 and α-2, creating a deeper active-site cleft. This additional loop is not essential in GST activity but it plays a role in substrate binding [46, 49, 53, 56, 74].

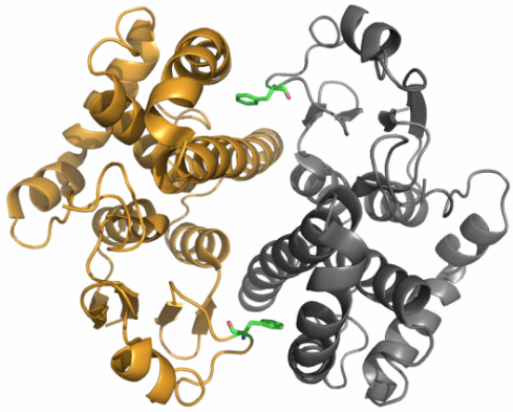


Figure 1.11. View of a class mu dimer (hGSTM2-2, PDB: 3GUR) down the two fold axis relating the two subunits, showing the 'lock and key' interaction of an aromatic residue (Phe-56) from domain 1 between α-4 and α-5 from domain 2 of the opposite subunit. This Figure was generated using PyMOL.

The γ-Glu moiety of GSH interacts with a hydrophilic complementary pocket near the subunit interface. The α and π class structures contain a salt bridge between an arginine (Arg-14 in Alpha, Arg-13 in Mu) and a glutamate (Glu-103 in Alpha, Glu-95 in Mu) which stabilizes the protein conformation at the γ-Glu site. A conserved tyrosine residue in β-1 (Tyr-8 in Alpha, Tyr-6 in Mu and Tyr-7 in Pi) has been proposed to activate the thiol group of GSH. Estimated pK_a values for the thiol group of glutathione are 5.7 – 6.9 for class mu, 6.7 – 7.0 for class alpha and 6.3 for class pi. In the alpha class structure, the sulfur atom of bound glutathione is also hydrogen-bonded to the guanidine group of Arg-14. In the mu class, GSH cysteinyl carboxyl group is not sequestered by

the protein backbone peptide loop connecting α-2 to β-3 as in the alpha and pi classes, but by a Trp-7. This disparity is responsible for the difference in GSH conformation taking place mostly at the Cys-Gly peptide bond in the mu class G-site, compared to the alpha and pi classes [46, 49, 53, 56].

In domain 2, only the interaction with a conserved aspartic residue of α-4 (Asp-100 in Alpha, Asp-105 in Mu and Asp-96 in Pi), which is part of the G-site, is common to all structures. The Alpha-class domain contains an extra C-terminal α-helix (α-9) which packs onto the hydrophobic site and contributes to a smaller and more hydrophobic active site. This extra helix is important for GST dimer stabilization and non-substrate ligand binding, and lowers both the rate of GSH binding and the pK_a value of the catalytic residue Tyr9 [46, 49, 53, 56, 75].

1.1.1.4.2.1.1.2. Theta-class

Theta-class GSTs are widespread in nature, they have been found in a diverse range of organisms including bacteria, plants and insects. Three main structural features distinguish this class from the Alpha, Mu and Pi classes, with which they only share 7% of similarity: theta class enzymes lack a "lock and key" interaction between their subunits; the cleft at their dimer interface is less pronounced; and their catalytic residue is a serine rather than a tyrosine. Substrate affinity has also been used to differentiate Alpha/Mu/Pi GSTs from those of the Theta class. Theta GSTs have the particularity not to bind affinity matrices such as GSH-agarose or *S*-hexyl-GSH-agarose, and most of them lack activity with CDNB. Two subunits (T1 and T2) were identified in human, sharing only 55% sequence identity and showing distinct substrate specificity. In particular, hGST T1-1 can conjugate GSH to halogenated organic compounds such as 1,2-epoxy-3-(*p*-nitrophenoxy)propane and shows some peroxidase activity with organic hydroperoxides; whereas hGST T2-2 is able to transfer GSH to sulfate esters, unlike other GSTs, and is believed to play an important role in the prevention of hepatocarcinogesis [53, 76-79].

1.1.1.4.2.1.1.3. Omega-class

A new class of GST was identified ten years ago by analysis of human expressed sequence tag sequences, and the crystal structure of hGST O1-1 was determined [80]. Since then, other omega GSTs have been found in varied organisms such as pig, rat, nematodes and insects. Omega class enzymes present several marked structural differences compared to the other classes: they have an unusual 19-20 proline-rich N-terminal extension; their domain 2 is composed of seven α-helices instead of the usual five; their inter-subunit interface is considerably more open; and their catalytic residue is a cysteine, unlike other eukaryote GSTs. GST O1-1 has low activity with most GST substrates, including CDNB but has a high thiol transferase activity. This isoform is strongly expressed in both skeletal and cardiac muscle and is thought to play a role in protecting cells from apoptosis by Ca^{2+} channel modulation [53, 80, 81].

1.1.1.4.2.1.1.4. Sigma-class

The Sigma class was identified by protein sequence alignments and crystal structural analysis revealed some important distinctive features, especially at the dimer interface and in the active site. Sigma class GSTs usually show hydrophilic interactions between the two monomers and lacks the 'lock and key' motif. Although GSH binds to the active site, Sigma class GSTs lack GST activity, which is thought to be due to insertion of eleven residues between the conserved α-4 and α-5 helices. This insertion closes up the conformation around the active site. Sigma GSTs also contain a third binding site for GSH conjugates which may play a role in conjugate transport [53, 82]. Hematopoietic Prostaglandin D-Synthases (H-PGDSs), which are involved in prostaglandin D (PGD) biosynthesis by catalysing the isomerisation of prostaglandin PGDH to PGD_2, belong to the GST Sigma class. Although they share more than 35 % of similarity with Sigma GSTs, H-PGDSs have a hydrophilic dimer interface with a lock and key interaction similar to that found in Alpha/Mu/Pi classes [83-85].

1.1.1.4.2.1.1.5. Zeta-class

The Zeta class was identified by phylogenetic analyses and was found in plants, worms and mammals. Zeta GSTs are also known as maleylacetoacetate isomerase (MAAI) and catalyse the glutathione-dependant cis-trans isomerisation of maleylacetoacetate to fumarylacetoacetate, which is the penultimate step in the catalysis of phenylalanine and tyrosine. MAAI is also involved in the detoxication of carcinogen dichloroacetic acid to glyoxylic acid. Zeta GSTs have low GSH-conjugating activity with CDNB. It can be explained by the presence of a very small, polar active site. Zeta class GSTs possess a hydrophobic dimer interface, lacking a 'V shape', as in the Theta class GSTs. There is a 'lock and key' motif where an unusual methionine residue plays the role of the key, and the catalytically essential residue is thought to be a serine (Ser14) [53, 86-88].

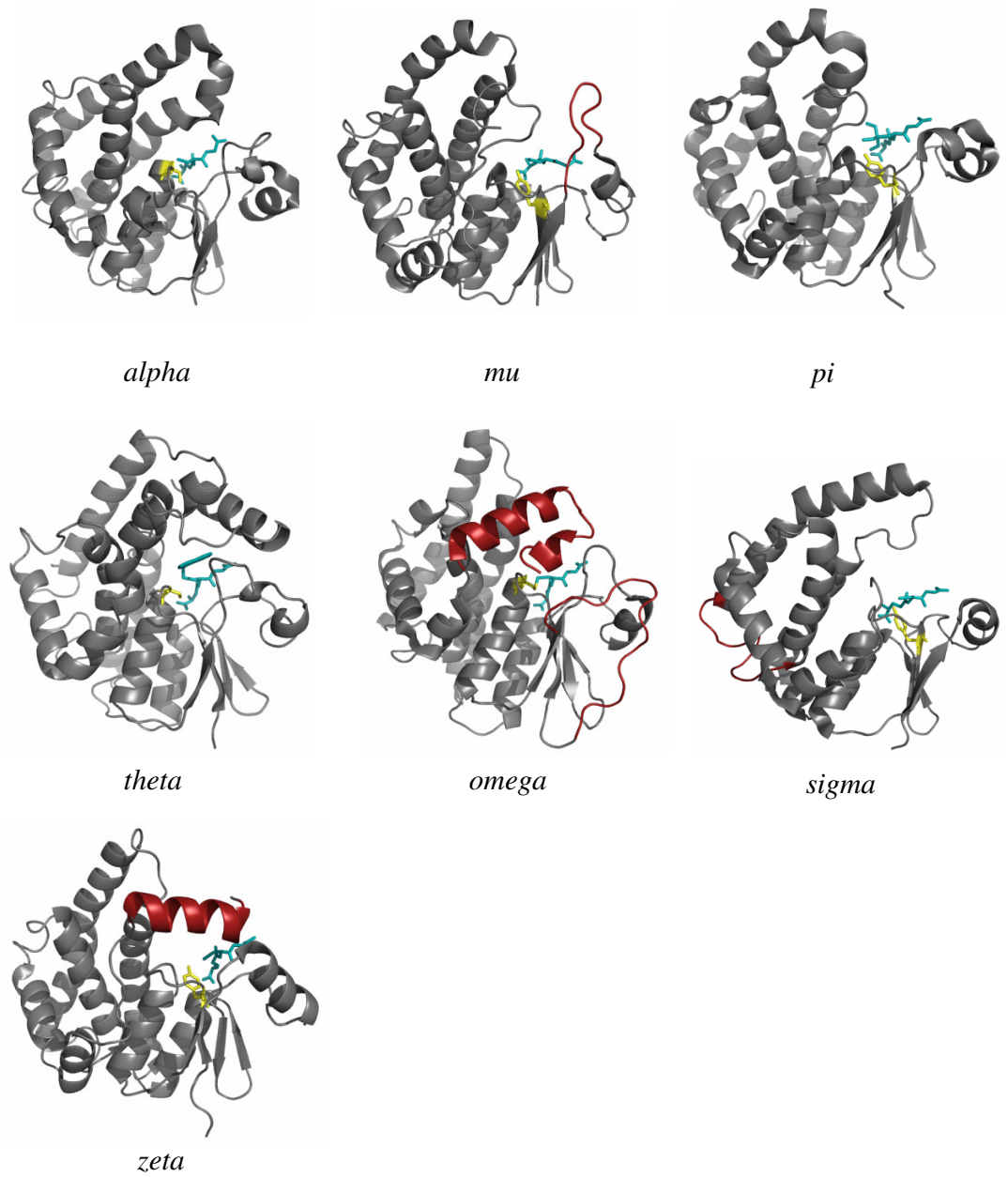


Figure 1.12. Crystal structures of *alpha* (hGST A2 complexed with GSH; PDB: 2WJU), *mu* (hGST M2 complexed with GSH; 3GUR), *pi* (hGST P1 complexed with nitrosyl glutathione; 2A2R), *theta* (hGST T1 complexed with 1-menaphthyl glutathione conjugate; 3LJR), *omega* (hGST O1 complexed with GSH; 1EEM), *sigma* (hGST S1 complexed with GSH; 3EE2) and *zeta* (hGST Z1 complexed with GSH; 1FW1). The views are of a single subunit perpendicular to the twofold axis of each dimer. Proteins are shown in grey, ligands are in blue, essential active residues (Tyr, Ser or Cys) are in yellow and isoform-specific structural features are coloured in red. This Figure was generated using PyMOL.

Table 1.1. Isoform-specific features of mammalian GSTs

GST class	*Number of isoforms*	*Essential catalytic residue*	*Structural/Activity specificity*
Alpha	5	Tyr/Arg	Extra α-helix (α-9)
Mu	6	Tyr	Mu loop between β-2 and α-2
Omega	1	Cys	Proline-rich N-terminal extension Open dimer interface Two extra α-helices in domain 2
Pi	1	Tyr	Unknown
Sigma	1	Tyr	Lack of a 'lock and key' motif at the dimer interface Hydrophilic dimer interactions 3rd binding site for GSH conjugates Prostaglandin D synthase activity
Theta	2	Ser	Lack of a 'lock and key' motif at the dimer interface Small active site Lack of a V-shaped dimer interface No binding to GSH affinity matrices
Zeta	1	Ser	Very small polar active site Lack of a V-shaped dimer interface Low GSH-conjugating activity with CDNB Maleylacetate isomerase activity Inhibited by dichloroacetic acid

1.1.1.4.2.1.2. Drug detoxification

Several anti-cancer agents such as busulfan, melphalan, chlorambucil or thiotepa are substrates of GSTs and can be directly inactivated through catalytic conjugation to GSH. For instance, GST-α from human colon adenocarcinoma cells are 40-fold more expressed in the presence of the anti-cancer drug chlorambucil (CMB), and catalyse its conjugation with GSH to form monochloromonoglutathionyl CMB. This conjugate is more water soluble and thus more easily removed from the cells (Figure 1.13.) [70, 73, 89-93].

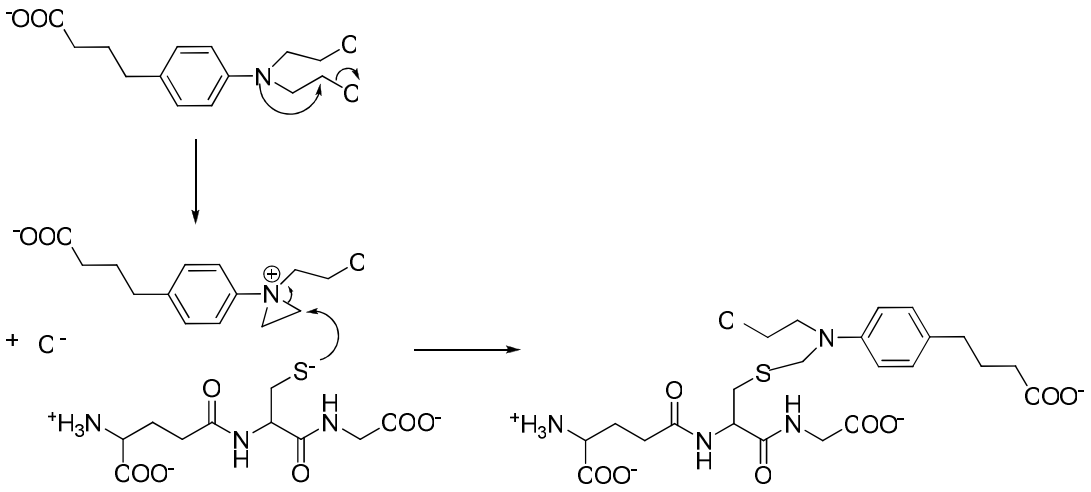


Figure 1.13. Conjugation of Chlorambucil and GSH via GST-α.

1.1.1.4.2.2. SjGST as a drug target in schistosomiasis treatment

1.1.1.4.2.2.1. Schistosomiasis

Schistosomiasis is a parasitic disease found in tropical and subtropical areas, caused by several species of platyhelminths of the genus *Schistosoma*. It is the second most serious parasitic disease after malaria. More than 200 million people are affected and around 200,000 deaths are caused annually. Three main species of flatworms are

responsible for schistosomiasis: *Schistosoma mansoni*, *Schistosoma japonicum* and *Schistosoma haematobium*. *S. mansoni* and *S. japonicum* cause the intestinal form of schistosomiasis, while *S. haematobium* causes the urinary form. Symptoms are varied depending on the species of infected schistosome but include abdominal pain, diarrhea, fever, haematuria and egg granulomas [94-96].

Resistance is developing to the leading anti-schistosomal drug – Praziquantel (PZQ; Figure 1.14.) – currently available for treatment. Although its mode of action is not exactly known at present, it is thought that PZQ increases the permeability of the parasite tegument for calcium ions. The drug thereby induces muscle contractions of the parasites resulting in paralysis in the contracted state [97, 98].

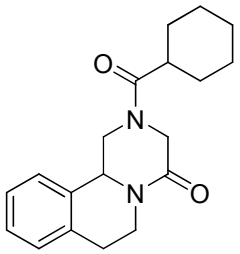


Figure 1.14. Structure of Praziquantel (PZQ).

1.1.1.4.2.2.2. The 26 kDa GST from *Schistosoma japonicum*

GSTs present a major defense mechanism for the schistosomes as they contain very low levels of other detoxification enzymes. Therefore, they are under investigation as targets for antischistosomal drugs to replace or to combine with those currently used. GSTs from helminthes are mostly expressed in the cytoplasm but secretory GSTs are also commonly expressed. The latter feature makes helminth GSTs tempting targets for vaccine development. *S. japonicum* possesses two GST isozymes, of 26 kDa and 28 kDa monomeric molecular weight. The 26 kDa enzyme has had great impact in biotechnology and is widely used as a fusion protein, allowing expression of proteins

which otherwise would have poor expression levels and simplifying purification through the use of immobilized glutathione [53, 95, 99-101].

The *S. japonicum* 26 kDa GST (SjGST) structure resembles the Mu-class GSTs. Domain 1 adopts the usual thioredoxin fold involving a four-stranded *β*-sheet surrounded by 3 *α*-helices. Between *β*-2 and *α*-2, sjGST shows a shortened version of the mu loop. Domain 2 contains five helices. There is a coil at the C-terminal end (residues 195-218) with a unique hairpin loop (211-218) exposed to the solvent. The C-terminal tail is close to the loop between β-2 and α2, helping to enclose the G site. As observed for the alpha, mu and pi mammalian isoforms described previously, the sjGST homodimer shows a 'lock and key' interaction. The essential catalytic residue is a tyrosine in position 7 (Figure 1.15.) [68, 102-104].

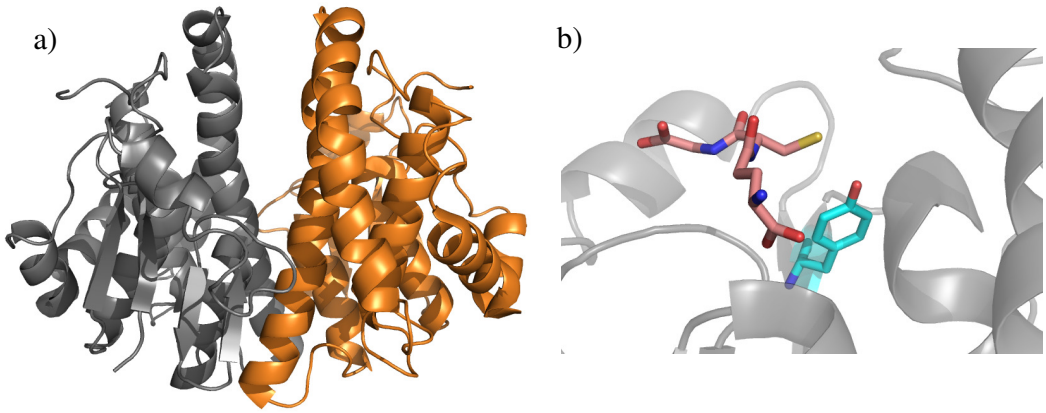


Figure 1.15. Crystal structure of homodimer (a: one monomer is shown in grey, the other monomer in orange) and G-site (b: catalytic tyrosine 7 is shown in blue, GSH in pink) of the 26 kDa SjGST in complex with GSH (PDB 1UA5). This Figure was generated using PyMOL.

To understand the mode of action of PZQ a crystal structure of the 26 kDa GST from *S. japonicum* bound to the drug has been determined. At therapeutic concentrations, PZQ binds one drug per GST homodimer in the intersubunit cleft adjoining the two catalytic sites (Figure 1.16.). Contacts are made with Gln67, Gly-97, Leu100, Asp101, Tyr104, and Arg108 from one subunit and with Tyr-104 of the second subunit. Most GSTs do not have the Tyr104 and are thus unlikely to bind the drug in the

same position. However, in contrast to its perceived mode of action, PZQ does not seem to inhibit sjGST enzyme activity. Its binding site is about 9.6 Å from the G site's essential tyrosine, so PZQ is not likely to bind with GSH [68, 105].

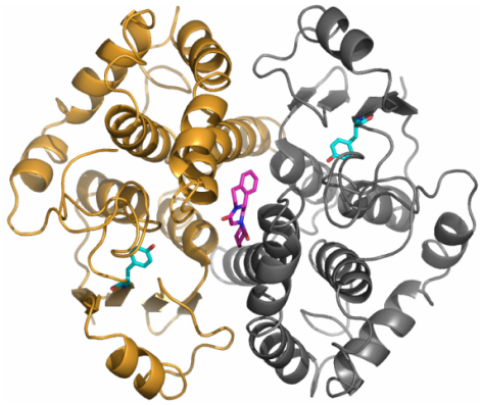


Figure 1.16. Crystal structure of the sjGST homodimer (one monomer in grey, the other monomer in yellow; PDB: 1GTB) showing PZQ (in pink) bound in the intersubunit cleft. The two catalytic tyrosine residues are shown in blue to indicate the positions of the active sites [68]. This Figure was generated using PyMOL.

1.1.1.4.2.3. Glutathione transferase inhibitors and pro-drugs

1.1.1.4.2.3.1. Inhibitors

Several approaches are currently used to synthesize GST inhibitors, including GS-R conjugates, GSH peptide analogs and nonpeptide analogs, bivalent inhibitors and ligandin-type inhibitors. A broad class of GST potential inhibitors exists, but they are therapeutically limited regarding their toxicity, carcinogenicity, or unsuitability as a drug for human use. Moreover, as GST up regulation in cancer cells often concerns certain isoforms, an ideal inhibitor should be isoform specific [44].

1.1.1.4.2.3.1.1. GS-R conjugates

The highly conserved G-site and the H-site versatility in GSTs have been used to design GSH-conjugate inhibitors. The GSH moiety allows anchoring of the ligand into the active site, whereas the variability of the H-site among the isoforms provides a source for isoform specificity. However, the fact that substrates can adopt different orientations within the H-site makes the drug rational design complicated.

An example of this is Ethacrynic acid (EA) a diuretic drug, commercially known as Edecrin, which is used to treat high blood pressure and the swelling caused by various medical problems such as congestive heart failure, liver failure and kidney failure. It acts by inhibiting a sodium-potassium-chloride transporter in the kidneys [106]. EA was also found to form a conjugate with GSH *via* Michael addition on its *α,β*-unsaturated ketone moiety, both spontaneously and by GST-driven catalysis. EA and its glutathione conjugate inhibit the GST-*α*, -*μ* and -*π* classes by direct binding to the isozymes active sites [107]. The crystal structure of human GSTP1-1 in complex with EA has been solved, showing that EA can bind GSTs in two different modes (Figure 1.17.) [108, 109]. After *in vitro* and animals studies, EA has been reported to potentiate the cytotoxic effects of anti-cancer agents such as chlorambucil and melphalan in human colon carcinoma cell lines [110].

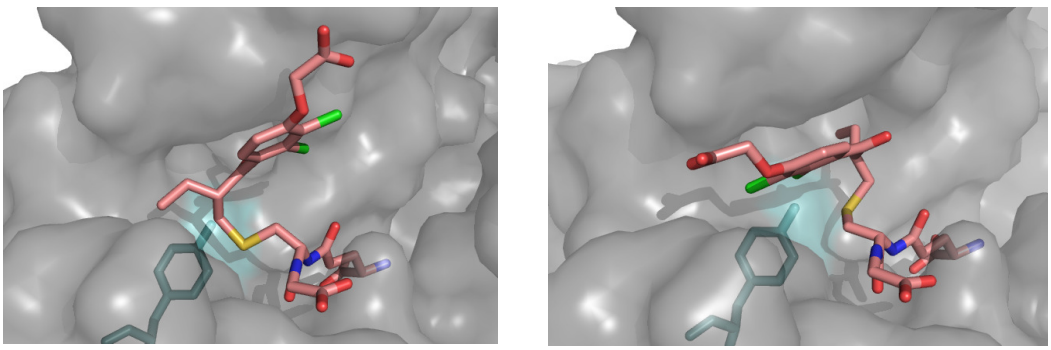


Figure 1.17. Crystal structures of hGSTP1-1 (PDB files: left 3GSS; right 11GS) showing the two binding modes of EA glutathione conjugate. hGSTP1-1 surface is shown in grey, EA conjugate in pink, the catalytic tyrosine residue in blue [108, 109]. This Figure was generated using PyMOL.

Other glutathione conjugates such as S-alkyl and S-benzyl GSH are used *in vitro* as probes, biochemical inhibitors or in affinity resins (Figure 1.18.) [111].

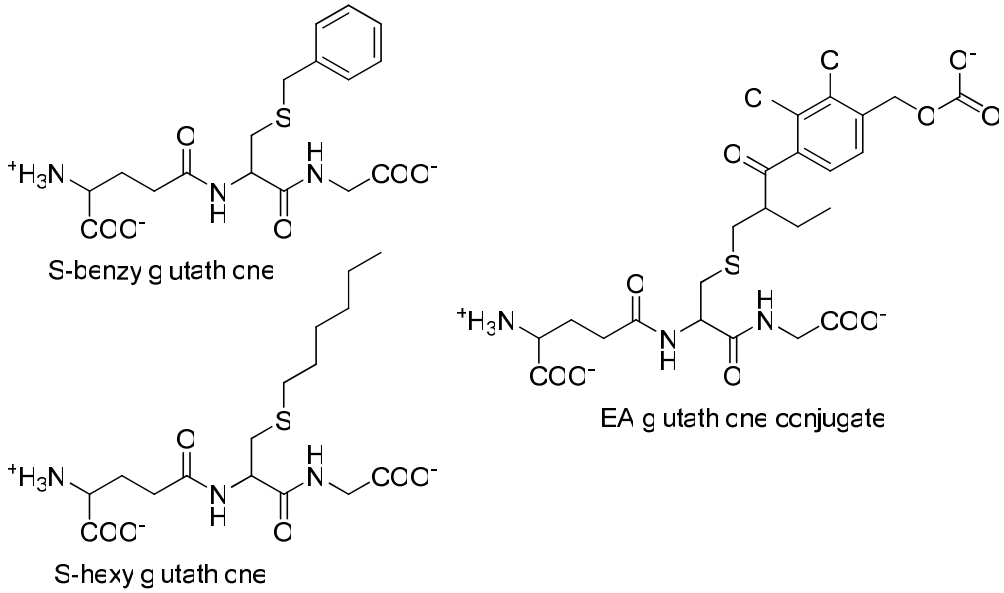


Figure 1.18. Chemical structures of EA, S-hexylglutathione and S-benzylglutathione.

1.1.1.4.2.3.1.2. Bivalent inhibitors

A new type of tight binding and isoform-selective inhibitors was developed by Atkins and co-workers in 2003 [112]. In their approach, the presence of two active sites separated by an accessible cleft in GSTs allowed them to design molecules with multiple binding domains. These inhibitors interact with active sites of each GST monomer simultaneously. As the distance between the actives sites vary among the isoforms, the length and nature of the linker connecting the inhibitor extremities interacting with the active sites allows for isoform selectivity [113]. Several bivalent compounds with different binding elements and linkers were synthesised. For instance, an inhibitor based on the EA-GSH conjugate was found to be 75 times more selective for GSTA1-1 over

GSTP1-1 (Figure 1.19.) [114]. The development of drugs based on this approach is currently being explored.

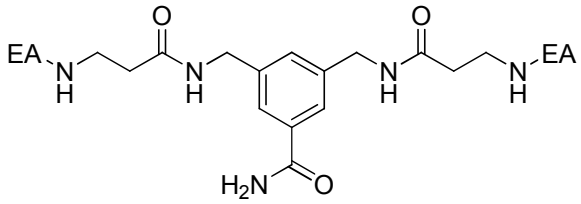


Figure 1.19. Bivalent inhibitor based on the GS-EA conjugate [114].

1.1.1.4.2.3.2. Glutathione transferase - activated pro-drugs

A pro-drug is a designed pharmacological substance which is administered in an inactive form and is activated by enzymes that are specifically elevated in target tissues. As GST levels are particularly high in cancer cells, they are good candidates for activating pro-drugs and releasing anti-cancer chemotherapeutic drugs in tumours. This approach allows an increased delivery of anticancer agents in a tumour tissue, and a minimized toxicity toward normal tissues [44, 73]. There are currently three types of GST targeted pro-drugs which are metabolized to turn into nitrogen mustards [115, 116], cytolytic nitric oxide [117] or thiopurine [118]. The most advanced of the GST-activated pro-drug candidates is TLK286, which is currently in phase 3 clinical trials. TLK286 is a GSH analogue sulfone derivative, activated by GST P1-1 and GST A1-1. The active site tyrosine would act as a base and promote a β-elimination reaction that cleaves the prodrug into a GST inhibitor (GS-vinyl-sulfone) and a DNA alkylating agent (nitrogen mustard) (Figure 1.20.) [44, 115, 116, 119].

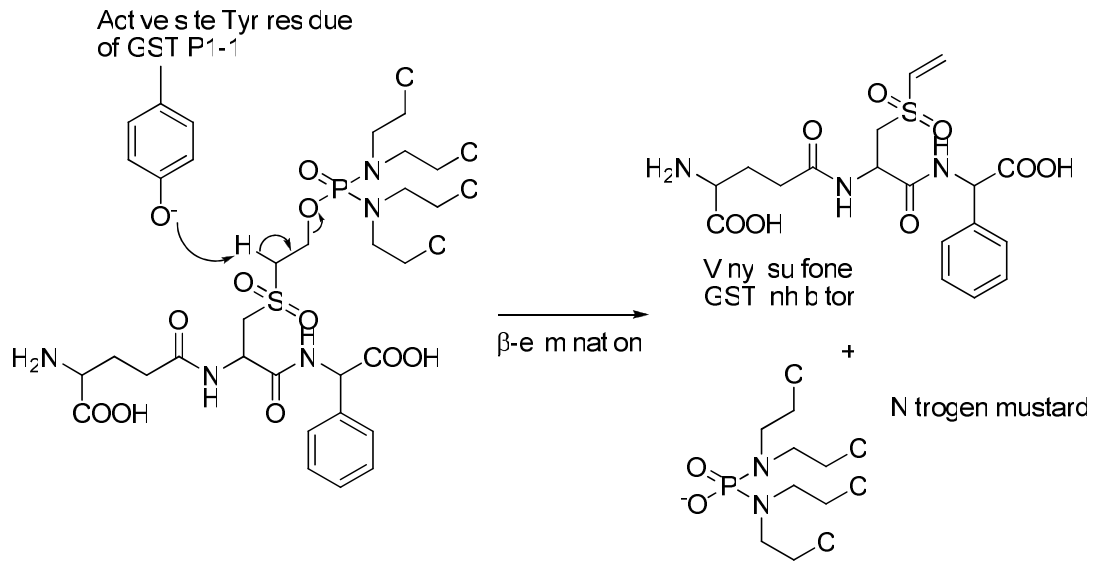


Figure 1.20. Activation of TLK286 (γ-glutamyl-α-amino-β-(2-ethyl-N,N,N',N'-tetrakis(2-chloroethyl)phosphorodiamidate)-sulfonyl-propionyl-(R)-(-) phenylglycine) by hGST P1-1.

1.1.2. Dynamic combinatorial chemistry as an exploring tool in isoform specificity

Dynamic combinatorial chemistry (DCC) is based on the reversibility of the reaction that links building blocks together. A dynamic combinatorial library (DCL) is governed by thermodynamics rather than kinetics and is therefore able to respond to external influences. According to Le Châtelier's principle: *'If a chemical system at equilibrium experiences a change in concentration, temperature, volume, or partial pressure, then the system alters in order to minimize, or counteract the effect of disturbance'*. Therefore, stabilisation of a library member through molecular recognition will generate a change in the equilibrium composition that will favour the formation of this species at the expense of the others.

1.1.2.1. Protein-directed dynamic combinatorial chemistry

Under thermodynamic conditions, a biological target such as an enzyme is used to select, stabilise and therefore amplify an active ligand or an inhibitor, which assembles within the confines of a binding site, directly from the library pool. The reaction between the components is then frozen, and the best binders identified and isolated through analysis of the DCL population (Figure 1.21.). Protein-directed DCC therefore constitutes a new efficient approach for studying, discovering, and ranking novel protein ligands, where chemical synthesis of drug candidates and biological binding assays are combined in only one step [120-124].

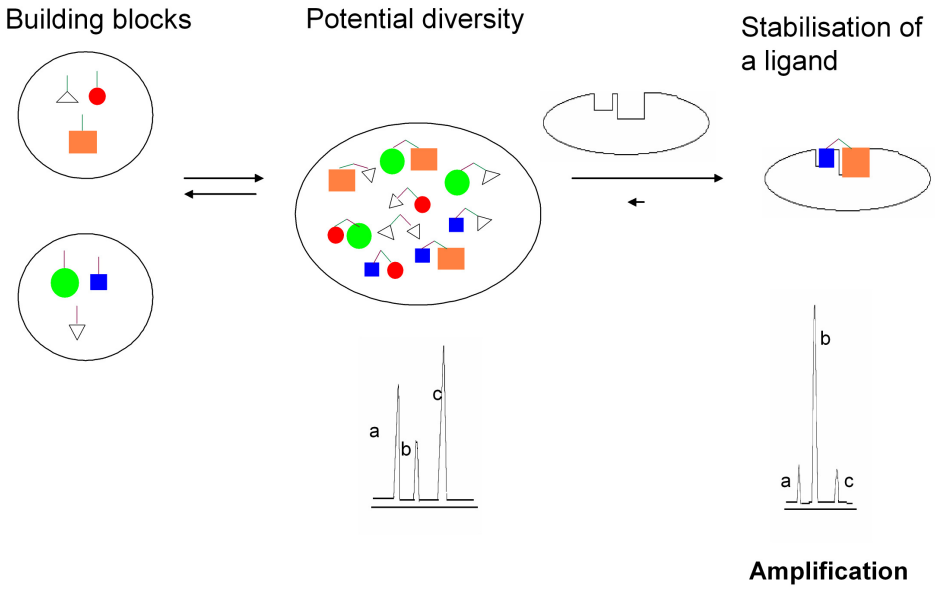


Figure 1.21. The principle of protein-directed DCC.

The utilisation of proteins requires the DCC to work at physiological conditions: pH, temperature and the nature of the reaction medium are three important parameters that have to be taken into account. Two typical DCC reactions fulfilling these conditions are the reversible imine and dithiol bond formations.

1.1.2.1.1. Imine bond formation

The use of imine bond formation in protein-directed DCC was first described by Lehn and Huc in 1997 [125]. In their study, they generated a DCL composed of imines to probe the active site of the enzyme Carbonic Anhydrase (CA), which is a drug target in the treatment of various diseases including neurological disorders and osteoporosis. The initial building blocks were composed of four amines and three aldehydes (Figure 1.22.) and were selected so that all the DCL components had structural features in common with known CA inhibitors. The functional groups creating the imine linkage had comparable reactivity. All components could be detectable by HPLC.

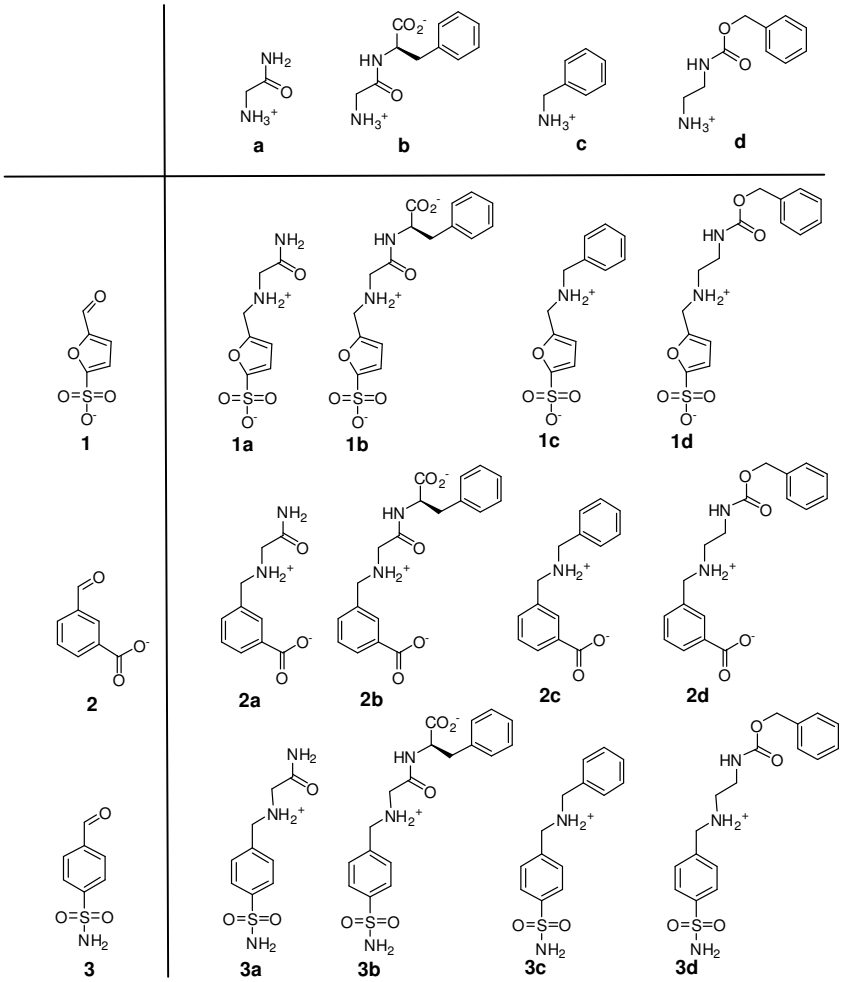


Figure 1.22. Lehn and Huc's imine DCL for CA inhibition.

Amines were in excess to create pseudo-first order behaviour with respect to the aldehydes, and to limit cross-reactivity between the aldehydes and any nucleophilic amino acid residues on the protein surface. The use of $NaBH_3CN$ then converted the imines into amines which are more stable and therefore easier to analyse by HPLC (Figure 1.23.).

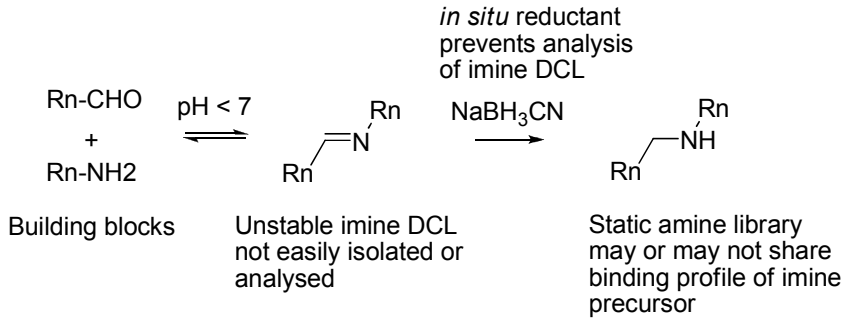


Figure 1.23. Reversible imine transamination reaction.

This experiment was the first proof of principle of DCC involving a biological target. However, there are two major problems associated with the utilisation of imines: first, because they are reduced to amines, imines are not represented in the final analysis; there is also a side reaction where starting aldehydes are reduced to benzylic alcohols, their concentration is therefore biased.

Lehn and co-workers also studied acylhydrazone formation (Figure 1.24.) as a method for DCL generation for inhibition of acetylcholine esterase (ACE) [126], which is a drug target in the treatment of glaucoma and Alzheimer's. The use of acylhydrazone formation in DCC in abiotic systems was introduced by Sanders and co-workers [127].

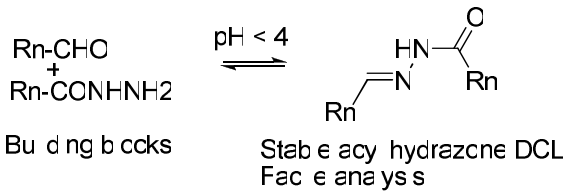


Figure 1.24. Reversible acyl hydrazone formation.

As the reaction takes place under conditions unsuitable for most enzymes and ACE in particular, the DCL, composed of thirteen building blocks (aldehydes and hydrazides, Figure 1.25.), was pre-equilibrated and frozen before ACE was added to the solution. A deconvolution approach was developed to identify the best ACE inhibitor. The pre-equilibrated 13-member DCL was frozen and assayed with ACE. Then, thirteen sub-libraries were prepared containing all building blocks minus one hydrazide or one aldehyde, frozen and assayed. The absence of the most active constituent would result in an increase in ACE activity. In parallel, each library component was re-synthesised and assayed individually. The results of these two experiments were in agreement, validating the DCC technique. The same approach was also used by Lehn and co-workers in a higher scale (440-member DCL) to identify a cationic inhibitor of a phosphotransfer protein kinase from *Bacillus subtilis* [128].

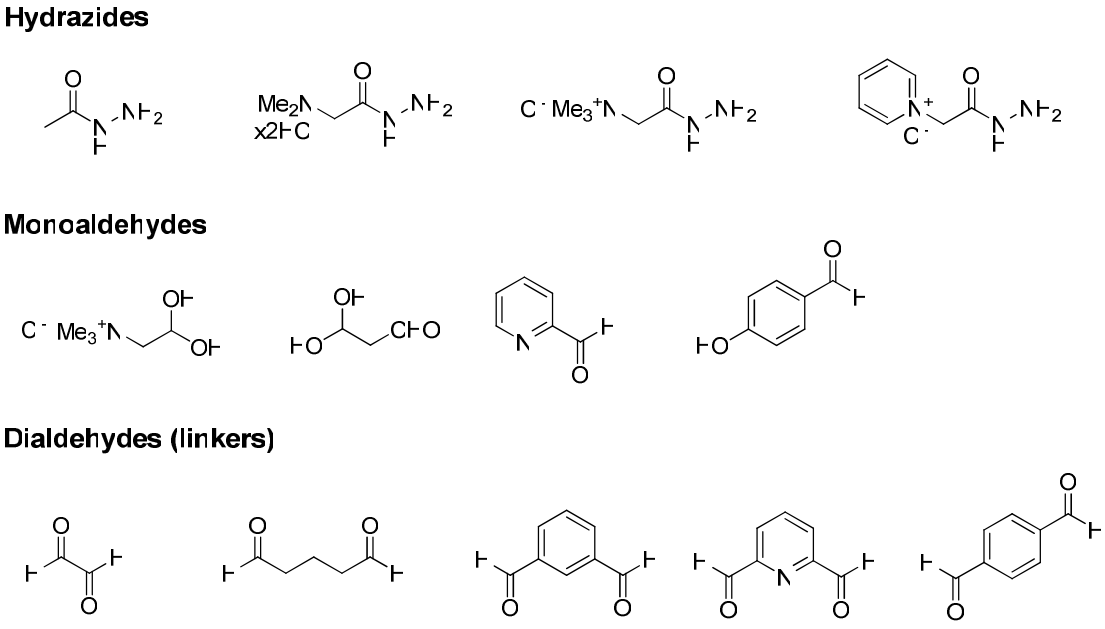


Figure 1.25. Hydrazide and aldehyde components for acyl hydrazone DCL

Following this proof of concept, the reversible acylhydrazone formation was applied to DCC by other groups, such as Eliseev and co-workers [129] and Beau and co-workers [130-132] who used this technique to generate new enzyme inhibitors.

1.1.2.1.2. Disulfide bond exchange

The disulfide bond exchange reaction was introduced to DCC in the late 1990s by Still, Sanders and Lehn in separate reports [133-135]. Given the fundamental role played by disulfide formation and cleavage in biology, this reaction is highly compatible with protein targets as disulfide exchange readily proceeds in water and at mildly basic pH. Unlike imines, disulfide bonds are stable enough to undergo HPLC analyses and isolation from aqueous solutions. For instance, Hunter and Waltho used disulfide exchange in peptidic DCLs to target the calcium transducer calmodulin (CaM) [136]. The DCL building blocks were composed of cystine dimers containing hydrophobic amino acid residues (Figure 1.26.) and the DCL was equilibrated for 48 hours. After equilibration with CaM, one dimer was significantly amplified and revealed to be a relatively good binder.

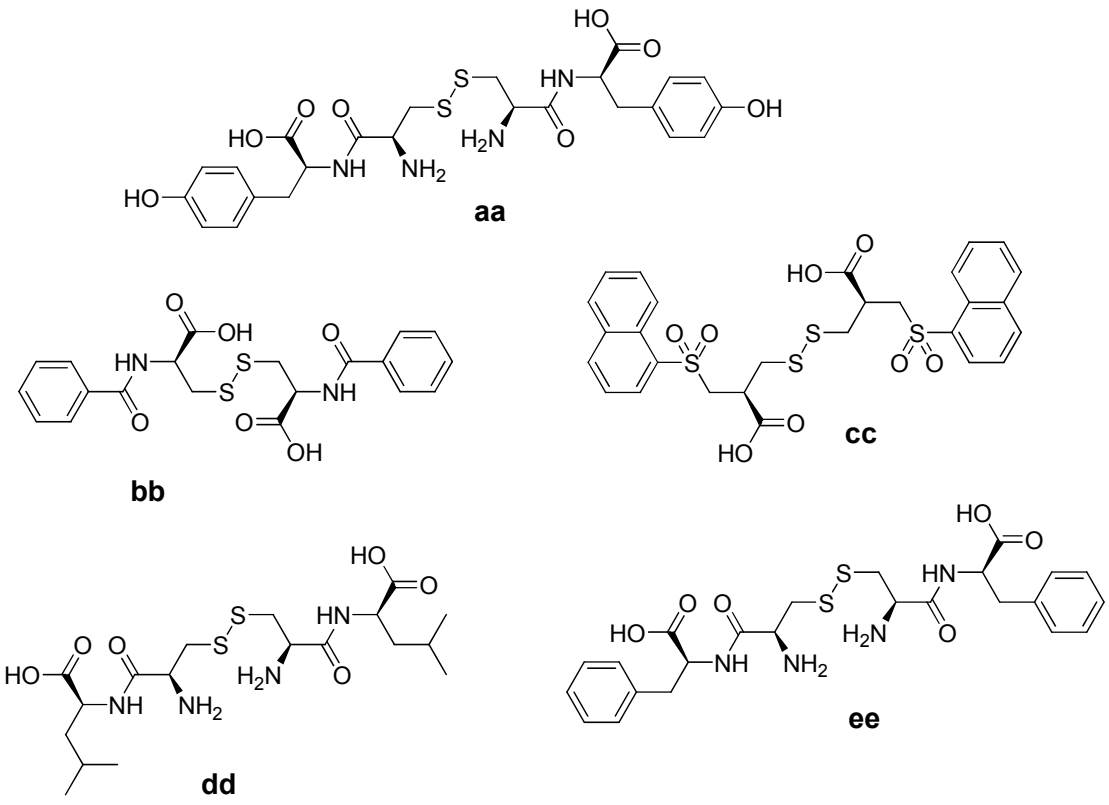


Figure 1.26. Cystine dimer components for calmodulin DCL.

1.1.2.1.3. Enzymatic methods

Enzymes have been used to catalyse reversible reactions in DCC. They present advantages such as being able to work within physiological conditions and can be applied to a variety of C-C and C-X bond-forming reactions. For instance, thermolysin was used in protease-catalysed amide-bond synthesis/hydrolysis [137-139].

Gleason and Kazlauskas introduced the concept of pseudo-dynamic combinatorial chemistry by combining compartmentalised irreversible synthesis of library with an irreversible destruction step [140, 141]. A pseudo-DCL of peptides was used to inhibit CA. The system was consisted of two dialysis bags suspended in a surrounding solution (Figure 1.27.). One of the bags was a synthesis chamber where dipeptide binders were irreversibly created through addition of amino acids to immobilised active esters. These dipeptides could diffuse to the solution – or screening chamber – containing the targeted CA and where a binding equilibrium was established. Dipeptides further diffused into the second dialysis tube – or destruction chamber – where they were hydrolysed back to amino-acid building blocks which could migrate to the synthesis chamber. After a number of cycles, this method proved to be highly selective.

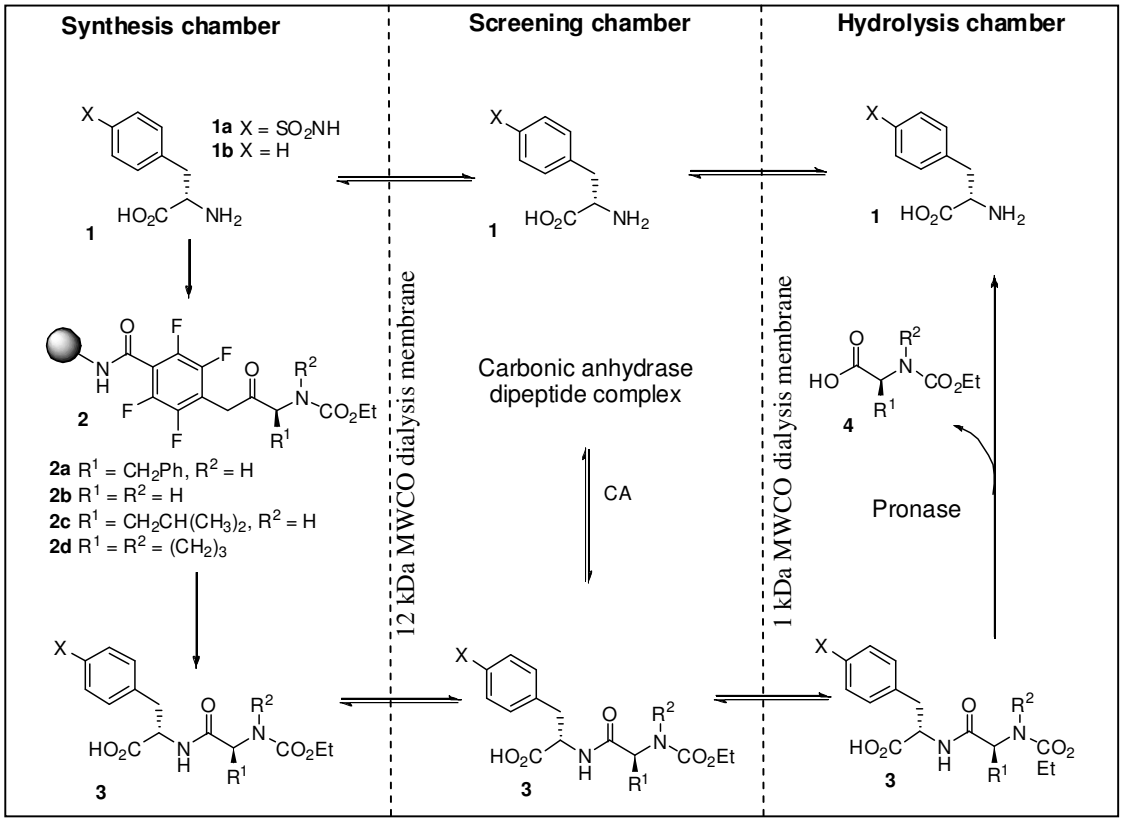


Figure 1.27. Enzymatic pseudo-DCL experiment.

1.1.2.2. Dynamic combinatorial chemistry and SjGST

In recent work by the Greaney and Campopiano groups, the 26 kDa GST from *S. japonicum* has been used as a template in DCC [142]. The reaction was based on the reversible conjugate addition of thiols to Michael acceptors, which is well suited for biological DCL synthesis as it takes place in water at room temperature and under mildly basic conditions [143]. Acidification easily switches the reaction off. A first DCL was constructed with GSH, three GSH analogues modified on the γ-glutamyl moiety and the Michael acceptor EA (Figure 1.28.). After one hour of equilibration, the library was analysed and the four EA adducts could be characterised by LC-MS. Upon addition of SjGST, the native substrate GS-EA conjugate (**7**) was highly amplified. Subsequent

binding assays, giving IC_{50} values of 0.32 μM for **7** and 88 μM for **10**, suggested a correlation between the observed amplification and binding affinities of the DCL components toward the enzyme.

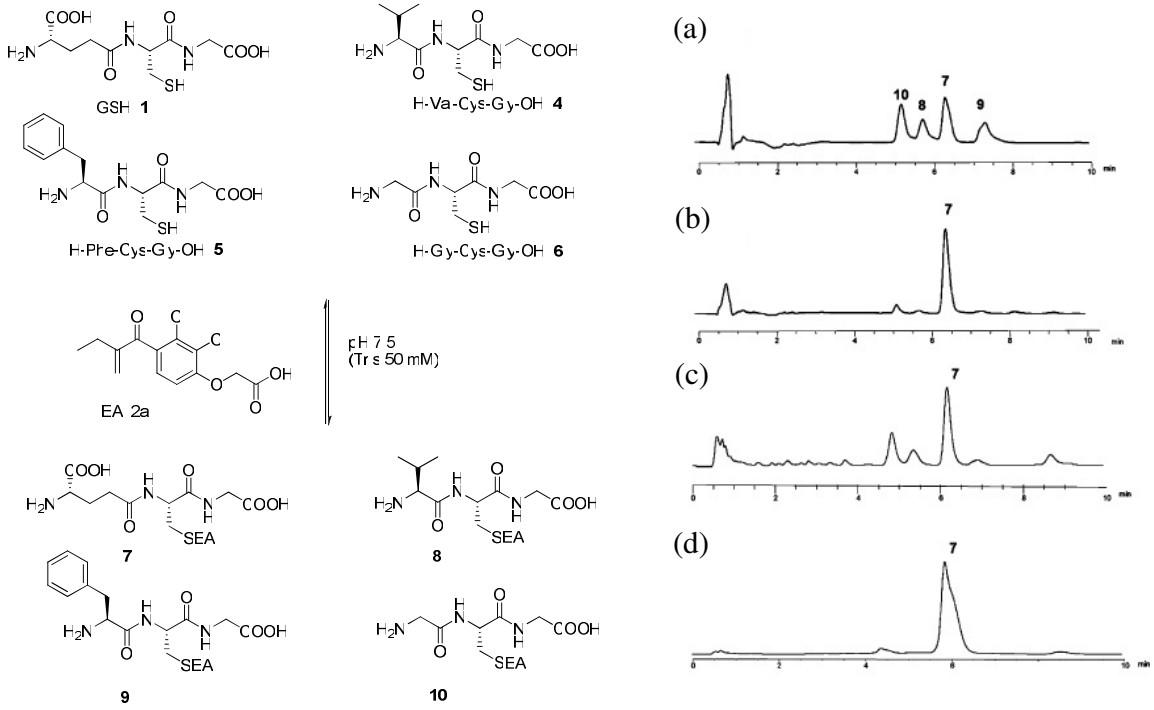


Figure 1.28. Thiol conjugate addition DCL template by sjGST: (a) DCL without enzyme, (b) DCL after 10 minutes in the presence of SjGST, (c) DCL after equilibration and subsequent addition of the target, after 2 days, (d) the same system after 6 days.

The work was extended to a larger library constituted of GSH and fourteen EA analogues (Figure 1.29.). In the presence of SjGST, three glutathione conjugates were amplified (**12a**, **12m**, **12n**) as they exhibited the strongest binding affinities for the enzyme. Both amplified GSH conjugates and their corresponding enone starting materials were assayed for inhibitory activity against SjGST and showed slightly higher IC_{50} values than GS-EA (IC_{50} = 0.32 μM).

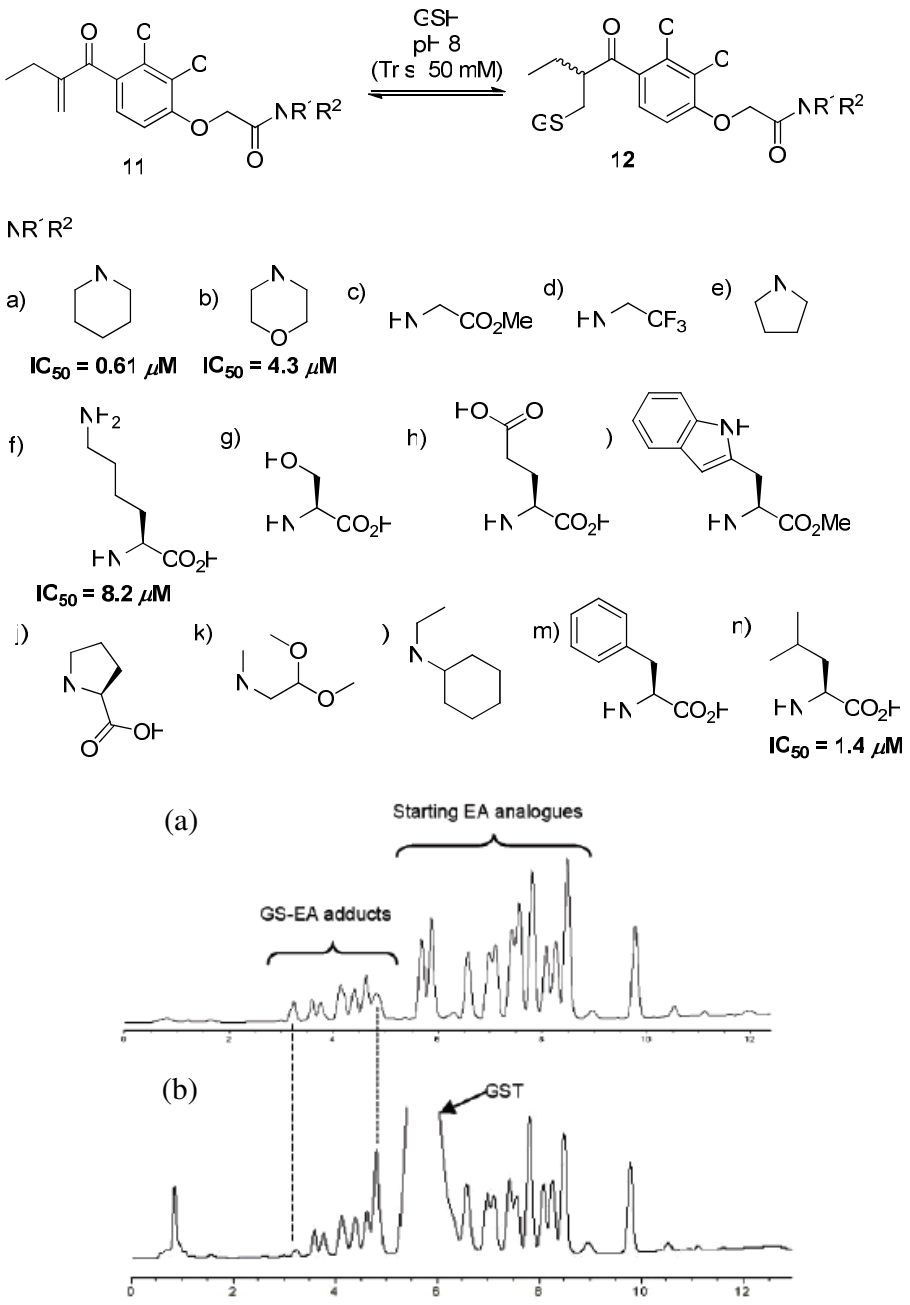


Figure 1.29. Thiol conjugate addition DCL templated by SjGST: (a) DCL without enzyme, (b) SjGST template DCL. IC_{50} values for **12**-**a**, **b**, **f** and **n** are indicated in bold.

1.1.3. Aims and research strategy

1.1.3. Aims and research strategy

The purification and biochemical characterization of the recombinant forms of four potential drug target GSTs from human (hGST P1-1), mouse (mGST A4-4; mGSTM1-1), and the parasitic worm *S. japonicum* (SjGST) were carried out prior to their application in a DCL study. The reversible acyl hydrazone formation, in the presence of aniline as nucleophilic catalyst, was used by Venugopal Bhat (University of Edinburgh, laboratory of Dr. Mike Greaney) to generate a DCL based on the CDNB scaffold in order to explore the active site of each purified GST. I consequently tested the DCL members in inhibition and binding studies with SjGST and hGST P1-1, in an attempt to correlate the amplification of a ligand and its affinity for the GST templating its molecular recognition.

1.2. Results and Discussion

1.2.1. Structural comparison of the four glutathione transferase isoforms

The four GST genes from the plasmids in our possession (pGEX-6P-1-SjGST, pET-15b-6His-mGSTM1, pET-15b-6His-hGSTP1 and pET-9a-mGSTA4) were all sequenced and aligned with the protein sequence database. The GST from pGEX6P-1 was identical to the *S. japonicum* 26 kDa GST (UniProt accession number: P08515). The two GSTs from the plasmids donated by Prof. John Hayes, University of Dundee – pET-9a-mGSTA4 and pET-15b-6His-mGSTM1 – were similar to GST A4 (gene ID: 14860, UniProt accession number: P24472) and GST M1 (gene ID: 14862, UniProt accession number: P10642) respectively, both isolated from *Mus musculus*. The GST from the plasmid given by Dr. Sylvie Blond – pET-15b-6His-hGSTP1 – corresponded to the human GST P1 (gene ID: 2950, UniProt accession number: P09211). SjGST from the commercial pGEX6P-1 plasmid contained an extra 26 amino acid sequence at the C-terminus, corresponding to the pre-scission protease and multiple cloning sites allowing the fusion of a protein. Analyses on the protein sequences revealed lengths varying between 222 and 244 amino acids and molecular weights of 23 to 28.5 kDa (Table 1.2.).

Table 1.2. Lengths and molecular weight of the four forms of GST

	Length (aa)	Molecular weight (Da)
SjGST	244	28,428
His_6-mGST M1	238	28,131
His_6-hGST P1	232	25,759
mGST A4	222	25,574

The protein sequences were then compared pair wise and the percentage of identity between each of them was determined. These data were incorporated into a matrix represented in Figure 1.30.

	SjGST	mGST M1	mGST A4	hGST P1
SjGST	100	44.2	28.4	29.1
mGST M1	44.2	100	22.3	31.7
mGST A4	28.4	22.3	100	32.6
hGST P1	29.1	31.7	32.6	100

Figure 1.30. Matrix representing the percentage of identity between each isozyme. Alignments were made with LALNVIEW, using a BLOSUM62 matrix [144].

The four isoforms presented low sequence similarities (from 22% to 44% identity). It is generally accepted that GSTs share greater than 40% identity within a class, and below 30% for GSTs belonging to different classes [53, 145, 146]; a low level of identity can be expected between homologous enzymes that catalyse a wide range of reactions on electrophilic substrates [146, 147]. Since SjGST is structurally very close to the mu class enzymes, it was not surprising to find the highest percentage of identity (44%) between SjGST and mGST M1-1 protein sequences [103, 104].

In order to identify the similarities between the GST isoforms, we aligned their ternary spatial conformations. Since the 3D-structure of mGST M1-1 has never been solved, a structural model was generated with the protein structure homology-modeling server SWISS-MODEL [148, 149]. The analysis was based on the high protein sequence resemblance with GST M1 from *Rattus norvegicus* (92.6% identity) [150], for which a macromolecular structure was already available in the Protein Data Bank (PDB file 6GSV_A). The backbone conformations of the monomeric isoforms SjGST [102], mGST M1, mGST A4 [151], and hGST P1 [152] were superimposed, and are illustrated in Figure 1.31. As reported by Dirr *et al.*[56], we observed that the four spatial structures follow a similar canonical fold despite a low sequence identity across the classes. Each monomer consists of two distinct domains connected through a small peptide linker. The

N-terminal domain 1, which constitutes roughly one third of the protein, consists of four β-sheets and three α-helices ($\beta\alpha\beta\alpha\beta\beta\alpha$), similar to that of the thioredoxin fold [54]. The four β-sheets are essentially in the same plane, two helices (α1 and α3) are below this plane and one (α2) is above it, exposed to the solvent. The C-terminal domain 2 is an all α-helical domain, generally consisting of five α-helices [46, 53, 56].

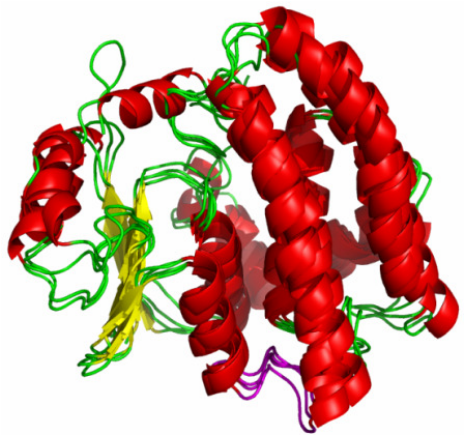


Figure 1.31. Ternary structural alignment of hGSTP1, mGSTA4, SjGST and the structural model of mGSTM1 generated with SWISS-MODEL. β-sheets are coloured in yellow, α-helices in red, inter-specific region in green and the inter-domain linker peptide in violet. This Figure was generated using PyMOL.

A protein sequence alignment based on both sequence homology and structural information was performed (Figure 1.32.). After highlighting the α-helices and β-sheets, we observed that the regions corresponding to the secondary-structure elements in the sequences of the GST enzymes could be aligned with confidence. Each subunit polypeptide chain folds into two structurally distinct domains that have a total secondary structure content of about 48-59% α-helix and 8-10% β-strands. The oligopeptide linking the two domains is composed of seven residues. Dirr *et al*. reported a number of 26 invariant residues between the alpha, mu and the pi classes, all located around the active site and mostly contributing to GSH binding, catalysis and conformational stability [56]. Similarly here, 22 conserved residues and 37 homologous residues were

found in the four isoforms, giving an overall homology of about 25% between the four sequences.

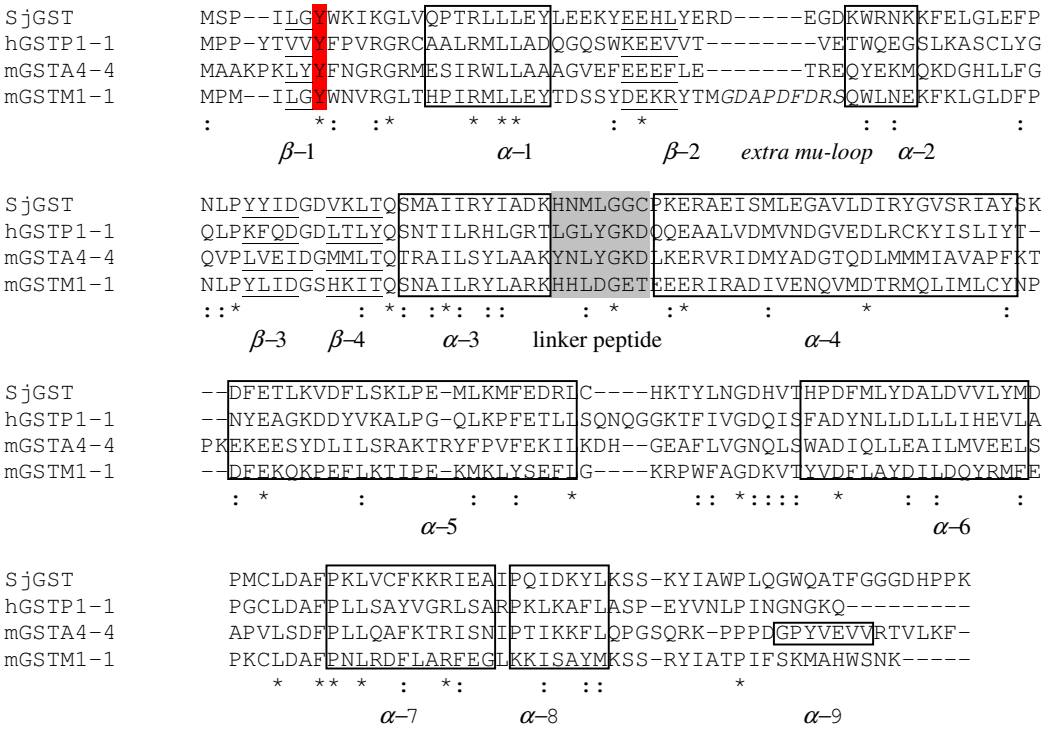


Figure 1.32. Alignment of topologically equivalent amino acid residues of the four GSTs. Catalytically active tyrosines are shown in red. Residues assigned to major secondary structure elements are underlined (beta strands) and squared (alpha helices). The seven-residue linker peptides are shown in grey. Invariant residues are indicated by an asterisk; residues with the same chemical properties are indicated by a double dot. α refers to an alpha helix and β to a beta strand. This alignment was generated with T-coffee [153].

Within the 22 conserved residues, three of them directly interact with GSH: Tyr7 (Sj) / 8 (P1) / 9 (A4) / 7 (M1) activates GSH and stabilizes the thionate anion; Gln66 (Sj) / 62 (P1) / 67 (A4) / 72 (M1) located between β–3 and α–3 and Asp100 (Sj) / 96(P1) / 101(A4) / 106(M1) both interact with the glutamyl moiety of GSH, their mutation provokes a considerable decrease in activity and GSH affinity [56]. Gly12 (Sj) / 13 (P1) / 14 (A4) / 12 (M1) and Pro55 (Sj) / 53 (P1) / 56 (A4) / 61 (M1) are both located in the active site and maintain a functional backbone conformation. Asp151 (Sj)

/ 152 (P1) / 157 (A4) / 157(M1) is located in α–6, in the hydrophobic pocket and plays a role in protein stability. Five leucine residues are involved in the hydrophobic core at the dimer interface: Leu19 (Sj) / 20 (P1) / 22 (A4) / 20 (M1); Leu20 (Sj) / 21 (P1) / 23 (A4) / 21 (M1); Leu136 (Sj) / 133 (P1) / 140 (A4) / 141(M1) makes Van der Waals contacts with Leu169 (Sj) / 170 (P1) / 175 (A4) / 175 (M1) and Leu175 (Sj) / 176 (P1) / 181 (A4) / 181 (M1). The twelve remaining conserved residues all play essential roles in protein conformation and stability.

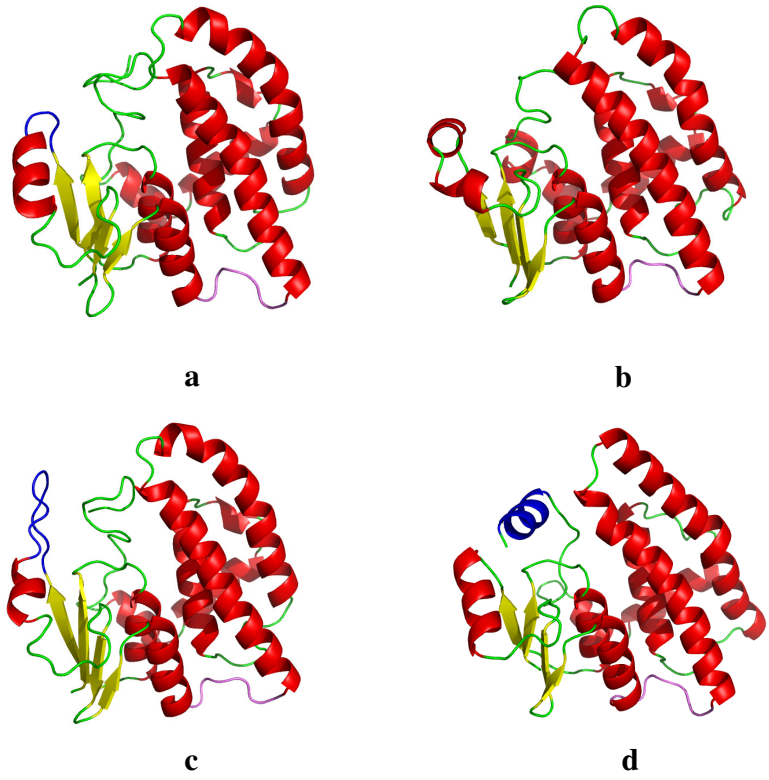


Figure 1.33. Three-dimensional structures of individual GST subunits. Helices are coloured in red, β-strands in yellow. The linker regions connecting the two domains are shown in violet. The extra mu-loop in mGST M1 and α-9 in mGST A4 are represented in blue. Protein database codes and references are given between parentheses: (**a**) SjGST (1Y6E, [102]); (**b**) hGST P1 (1EOG, [152)]; (**c**) mGST M1; (**d**) mGST A4 (1GUK, [151]).

Despite the similarities between the protein sequences, comparing each individual three-dimensional structure allowed us to point out some obvious differences in their backbones (Figure 1.33.). Each of mGST M1-1 and mGST A4-4 presents a unique structural feature, typical from the mu and the alpha class, respectively. M1-1 has a characteristic "mu loop" which is located between β–2 and α–2. A4-4 N-terminus has an extra α-helix (α–9), which packs on to the H-site. Both of these specific structural elements are located close to the substrate binding sites and contribute to a more constricted and less solvent-exposed active site of these two isoforms compared to the other GSTs. SjGST contains a smaller version of the "mu loop" observed in mGST M1, which is not long enough to cover the binding sites.

Differences in H-site topologies are correlated to the distinct substrate specificities observed in the GST family. As well as structural characteristics, kinetic properties constitute another important criterion in isozyme differentiation. I expressed, purified and kinetically characterised the four isoforms, using the GST-universal CDNB assay. A catalytically inactive mutant of SjGST – Y7F-SjGST – used later on as a control in the DCC experiments carried out by Venughopal Bhat, was also generated and characterised.

1.2.2. Glutathione transferase purification and characterization

1.2.2.1. Expression and purification

SjGST, mGST A4-4, His_6-mGST M1-1 and His_6-hGST P1-1 proteins were expressed at 37°C in *E. coli* BL21 (DE3) host cells using prokaryotic expression vectors pET or pGEX. After induction with isopropyl β-D-1-thiogalactopyranoside (IPTG) for 3 to 5 hours, an obvious increase in expression of the recombinant proteins of the expected mass (~26 to 28 kDa) was observed using SDS-PAGE (Figure 1.34.).

The conserved Tyr7 active site residue is known to play a critical role in GSH conjugation for SjGST, stabilizing the GSH thiolate anion through H-bonding from the phenol group, with enzymes lacking this residue being catalytically inactive. I prepared a SjGST Y7F mutant form, in which the crucial tyrosine residue is replaced with

phenylalanine. pGEX-6P1 was used as a template for mutagenesis and the mutation was confirmed by DNA sequencing. Y7F SjGST was expressed in *E. coli* BL21 at 30˚C and induced with IPTG for 5 to 6 hours (Figure 1.34.).

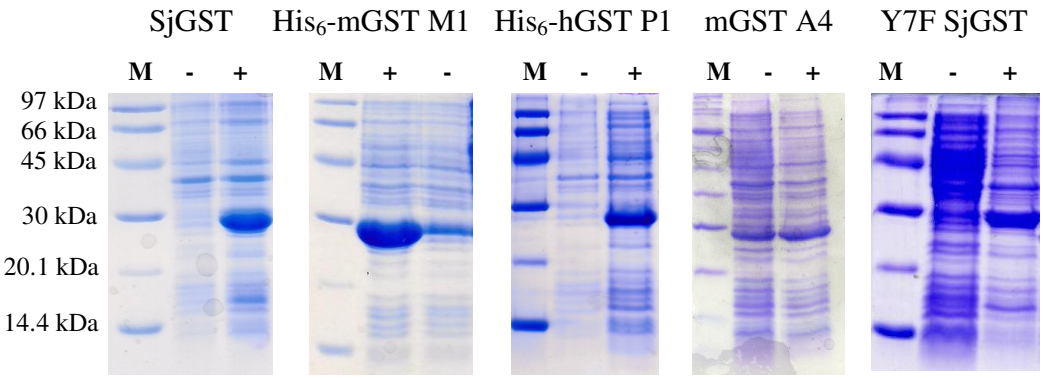


Figure 1.34. SDS-PAGE showing expression of proteins from *E. coli* before (-) and after (+) induction with IPTG. From comparison with the marker (M) bands it is possible to see a large increase in the concentration of proteins of approximately 26 kDa.

All the plasmids, except pET9-a-mGST A4, contain a *lacI* gene so the recombinant proteins from these vectors should only be expressed in the presence of IPTG. The lac I repressor may not bind tightly enough to the lac O operator of pET15-b-His_6-mGST M1, allowing a slight expression of His_6-mGST M1 without inducer.

Purifications of SjGST, mGST A4-4 and mutant Y7F SjGST were performed by GSH affinity chromatography, using a 20 ml GSTPrep FF 16/10 column, followed by a gel-filtration chromatography using a 320 ml Sephacryl S-200 column. Proteins with a histidine tag – His_6-hGST P1-1 and His_6-mGST M1-1 – were isolated using a 5 ml Nickel column and a 320 ml Superdex S200 size exclusion column. Cell free extracts, affinity column flow-throughs (containing unbound proteins), washes and the eluted fractions were analysed by SDS-PAGE (Figures 1.35. to 1.39.). Specific activities were measured at each purification step (Tables 1.3. to 1.6.). The catalytic activity of the SjGST_Y7F mutant was also checked after purification and compared to the wild type SjGST (Figure 1.40). Kinetic parameters of each purified GST are presented and discussed in section 1.2.2.2.

Purification and characterization of SjGST

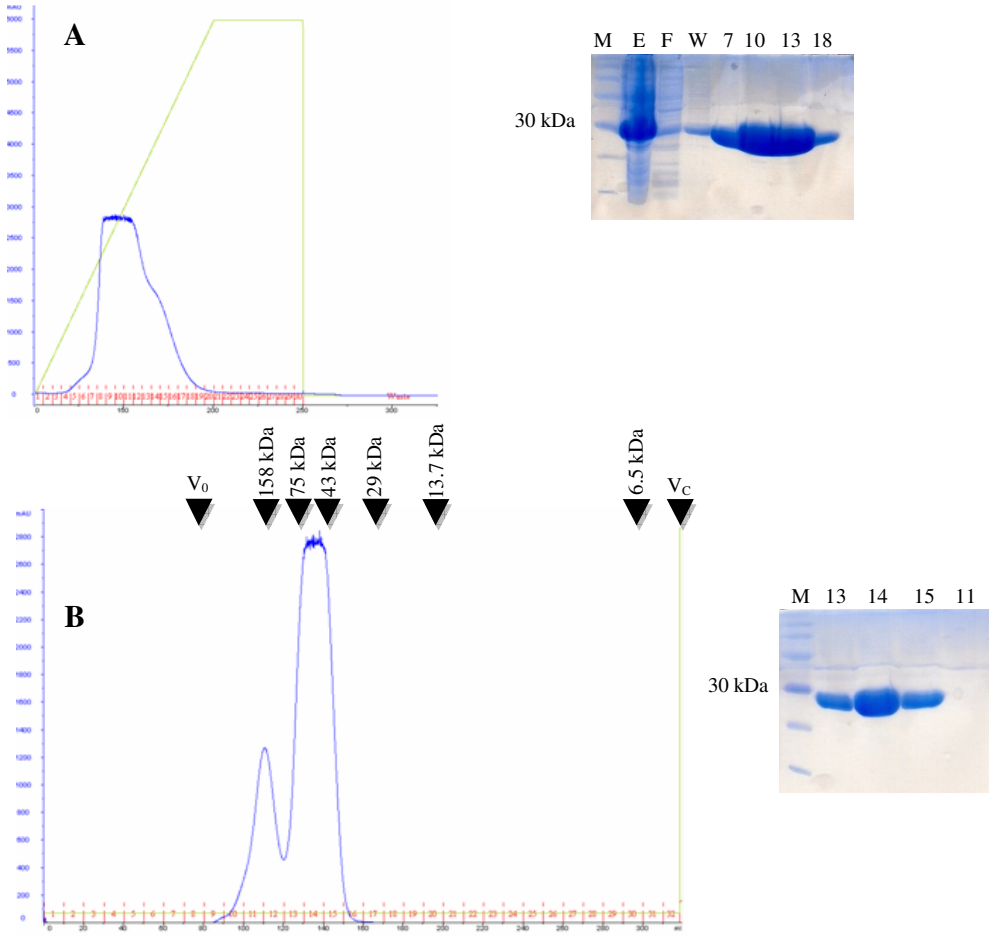


Figure 1.35. Elution profiles of sjGST from the GSTPrep column (**A**) and from the Sephacryl S-200 (**B**) showing the A_{280}. Calibration markers are specified at the top of the gel filtration S-200 column UV trace, V_0 is a void volume, V_C is the column volume. Collected fractions were analysed by SDS-PAGE (M: low molecular weight marker; E: cell free extract; F: flow through; W: wash; collected fractions are represented by their number).

Table 1.3. Purification of SjGST

	Volume (ml)	Total protein (mg)	[Protein] (mg/ml)	Total activity (µmol/min/ml)	Specific activity (µmol/min/mg)	Yield (%)
Cell free extract	50	371	7.4	501.2 ± 59	67.7 ± 8	100
S-200 eluate	30	91	3	315.3 ± 51	105.1 ± 17	63

Purification and characterization of mGST A4-4

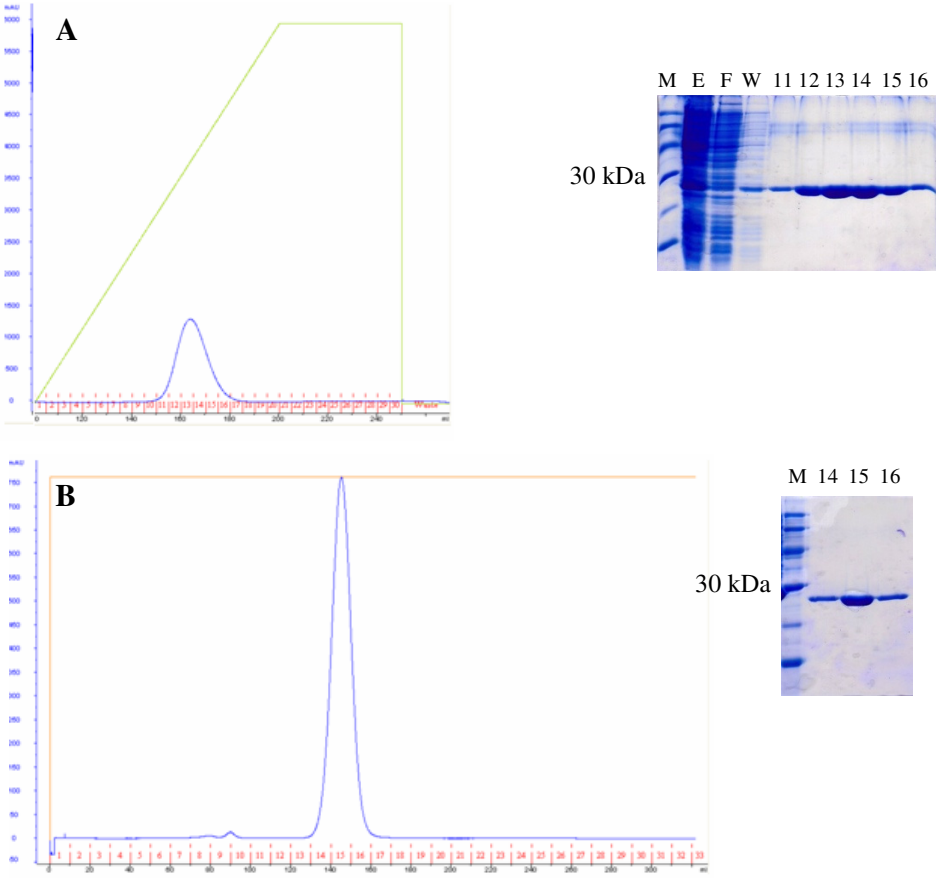


Figure 1.34. Elution profiles of mGST A4-4 from the GSTPrep column (**A**) and from the Sephacryl S-200 column (**B**) showing the A_{280} and SDS-PAGE analyses of the collected fractions.

Table 1.4. Purification of mGSTA4-4

	Volume (ml)	Total protein (mg)	[Protein] (mg/ml)	Total activity (µmol/min/ml)	Specific activity (µmol/min/mg)	Yield (%)
Soluble cell lysate	50	295	5.9	20.3 ± 1.9	3.4 ± 0.3	100
S-200 eluate	30	23	0.8	11.9 ± 1.2	14.9 ± 1.5	59

Purification and characterization of His_6-mGST M1-1

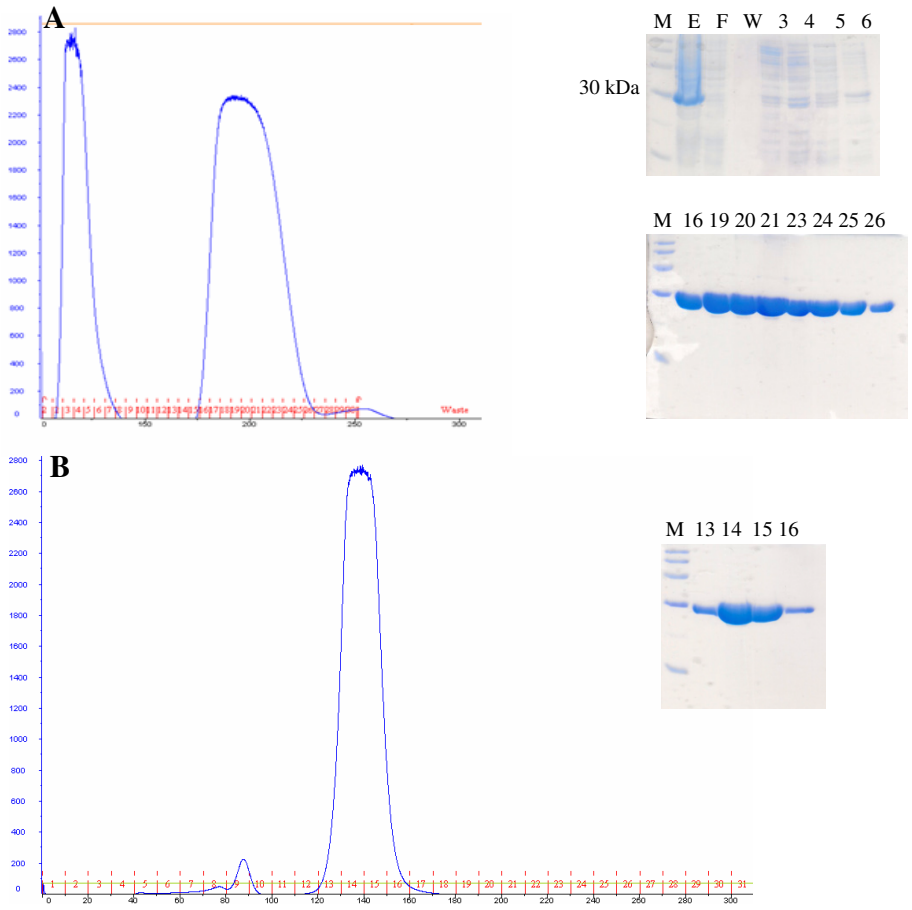


Figure 1.37. Elution profiles of His_6-mGST M1-1 from the Nickel column (**A**) and from the Sephacryl S-200 column (**B**) showing the A_{280} and SDS-PAGE analyses of the collected fractions.

Table 1.5. Purification of His_6-mGST M1-1

	Volume (ml)	Total protein (mg)	[Protein] (mg/ml)	Total activity (µmol/min/ml)	Specific activity (µmol/min/mg)	Yield (%)
Soluble cell lysate	50	278	5.6	58.9 ± 0.6	10.5 ± 0.1	100
S-200 eluate	40	61	1.5	40.8 ± 0.5	27.2 ± 0.3	69

Purification and characterization of His_6-hGST P1-1

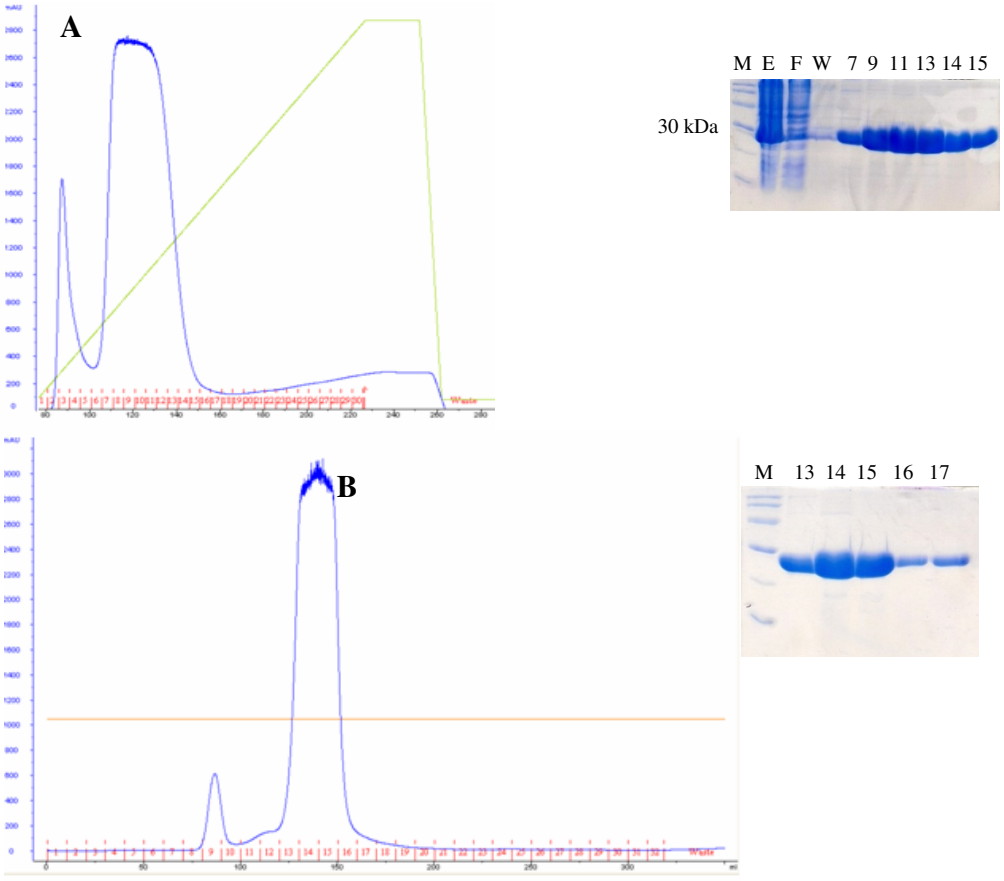


Figure 1.38. Elution profiles of His_6-hGST P1-1 from the Nickel column (**A**) and from the Sephacryl S-200 (**B**) showing the A_{280} and SDS-PAGE analyses of the collected fractions.

Table 1.6. Purification of His_6-hGST P1-1

	Volume (ml)	Total protein (mg)	[Protein] (mg/ml)	Total activity (µmol/min/ml)	Specific activity (µmol/min/mg)	Yield (%)
Soluble cell lysate	50	284	5.7	353.1 ± 9.9	61.9 ± 1.7	100
S-200 eluate	40	76	1.9	260.9 ± 7.4	137.3 ± 3.9	74

Purification and characterization of Y7F SjGST

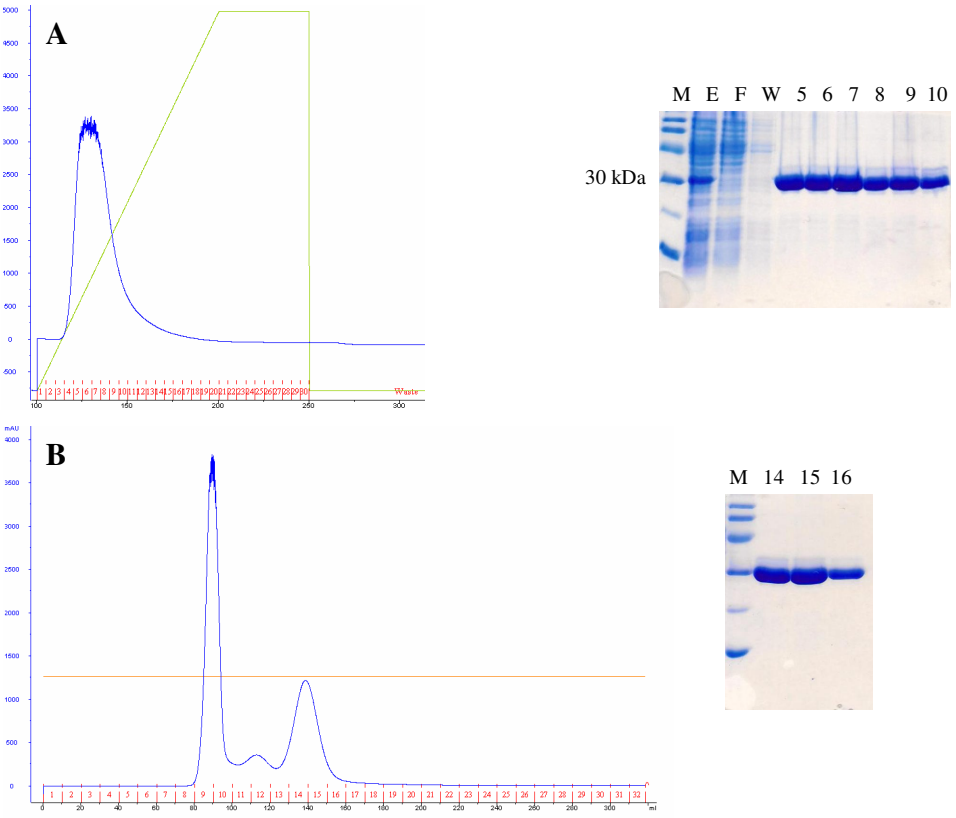


Figure 1.39. Elution profiles of Y7F from the GSTPrep column (**A**) and from the Sephacryl S-200 (**B**) showing the A_{280}. The collected fractions were analysed by SDS-PAGE.

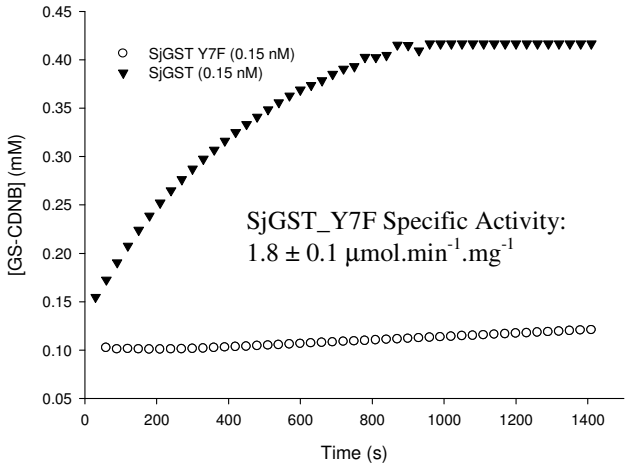


Figure 1.40. Comparative CDNB assays between SjGST WT and SjGST_Y7F, the mutant catalytic activity was reduced to 1.7%.

A yield of ~20 mg (mGST A4-4) to ~90 mg (SjGST) of recombinant protein was generally obtained from three liters of *E. coli* BL21 (DE3) culture. The SjGST_Y7F mutant was isolated in a yield of ~25 mg.

The five GSTs eluted as homodimers from the gel filtration column in agreement with their predicted dimeric structures [103, 108, 151]. They were purified with total recoveries of about 60 and 70% of the total activity using CDNB as a substrate in a standard GST assay, for GSTs purified with GSH affinity chromatography (mGST A4-4 and SjGST) and GSTs purified with Ni^{2+} affinity chromatography (His_6-mGST M1-1 and His_6-hGST P1-1), respectively. In a previous study, Blond and co-workers expressed the His_6-hGST P1-1 using the same plasmid, the same *E. coli* strain and a Co^{2+} affinity chromatography, with a total recovery of about 67% from the cell free extract [154].

In order to identify the purity of each sample and to determine the mass of each purified GST, ESI LC-MS analysis was carried out. A preliminary LC-MS of SjGST after dialysis against 0.1 M KPhos buffer, pH 6.8 gave a result of 29,350 Da (Figure 1.41. A).

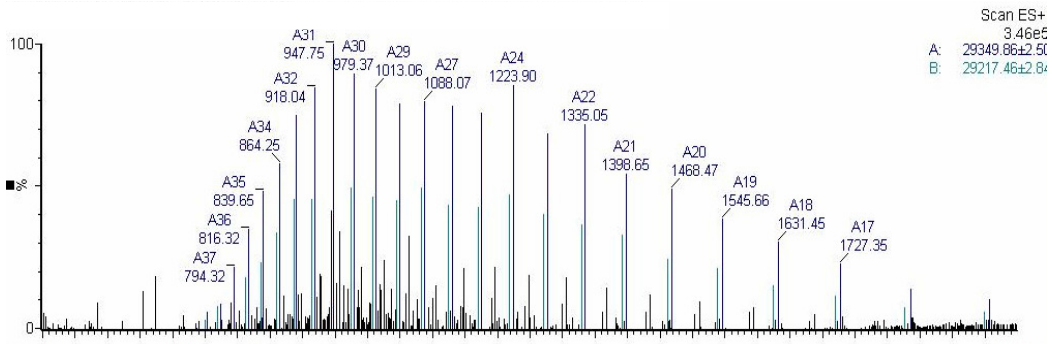


Figure 1.41. The mass spectrum of the purified sjGST. The ion envelope shows peaks ranging from 750 to 2000. The deconvoluted mass was determined using MassLynx.

Two ion series was observed in the spectrum, the first ion corresponds to the theoretical mass (28,428 Da) of the monomeric protein bound to three molecules of glutathione (3 x 307 Da). Another set of peaks (Figure 1.41. B) indicated a mass of 29,217 Da, which corresponds to the SjGST monomer lacking a methionine (132 Da). This suggests that a part of the proteins had their N-terminal methionine cleaved during

their synthesis. In order to obtain a ligand-free protein, the sample was dialysed against 4 L of KH_2PO_4 buffer, containing 1.4 mM of β-mercaptoethanol allowing the disulfide bonds to be reduced. LC-MS analysis of this sample gave results of 28,434 Da (Figure 1.42. A) and 28,303 Da (Figure 1.42. B) which are in agreement with the theoretical masses of the monomer and the N-terminal truncated monomer respectively. Looking at SjGST's crystal structure, I found that each monomer possesses four cysteines (Cys83, Cys135, Cys166 and Cys175) which are all solvent accessible [102]. These results suggest that three of four cysteines can bind to GSH.

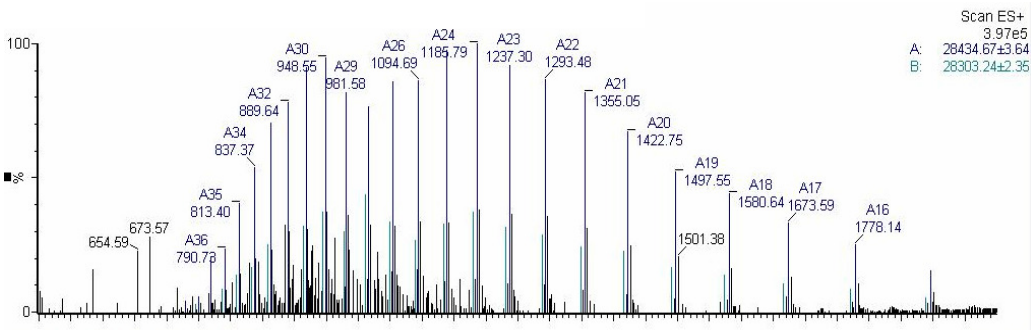


Figure 1.42. Mass spectrum of the purified SjGST after dialysis with KPhos buffer and *β*-mercaptoethanol. The ion envelope shows peaks ranging from 650 to 2000. The deconvoluted mass was determined using MassLynx.

mGST A4-4 does not contain any cysteine residues and therefore did not require any reducing agent in the dialysis buffer. LC-MS analysis on mGST A4-4 gave a result of 25,450 Da (Figure 1.43.), which is in agreement with the theoretical mass of the monomer (25,445 Da).

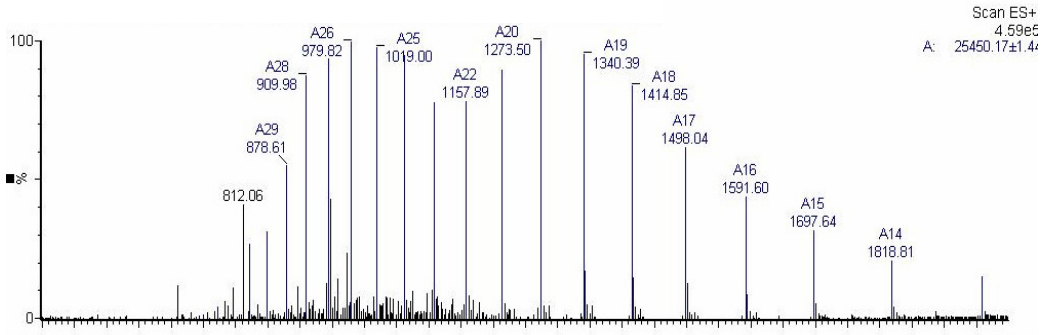


Figure 1.43. Mass spectrum of the purified mGST A4-4. The ion envelope shows peaks ranging from 750 to 2000. The deconvoluted mass was determined using MassLynx.

His_6-hGST P1-1 and His_6-mGST M1-1 were dialysed against KPhos and Tris/HCl buffers respectively, prior to MS analysis. A mass of 25,630.6 Da was obtained for His_6-hGST P1-1 and is in agreement with the calculated theoretical mass of 25,630 Da (Figure 1.44.). A MALDI-TOF analysis on His_6-mGST M1-1 gave a major peak with a mass of 28,130.47 m/z, which corresponds to the theoretical mass of 28,131 Da (Figure 1.45.). A second peak, lower in intensity, was observed at 56,256 m/z and matches with the theoretical mass of the dimeric form of His_6-mGST M1-1 (56,262 Da).

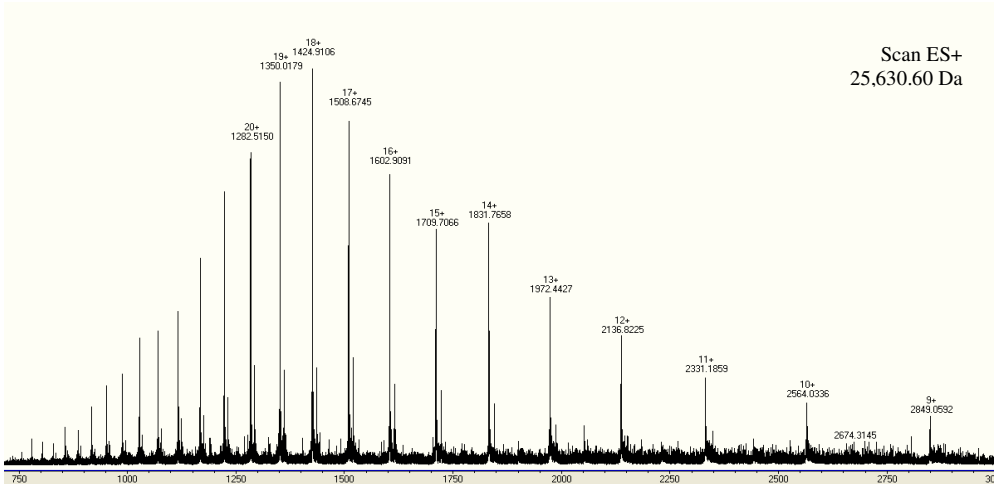


Figure 1.44. LC ESI-MS spectrum of the purified His_6-hGST P1-1. The ion envelope shows peaks ranging from 750 to 3000. The deconvoluted mass was determined using MassLynx.

The purified mutant was dialysed against Tris/HCl (20 mM) prior to MALDI-TOF mass spectrometry analyses. As for SjGST, β-mercaptoethanol was added to the dialysis buffer to reduce any eventual disulfide bonds with GSH. The analysis gave a mass of 28,410.54 m/z, corresponding to the theoretical mass of SjGST_Y7F (28,412 Da) (Figure 1.46.). A second peak with a mass of 56,821.94 m/z was observed and corresponds to the dimeric form of the mutant (56,824 Da).

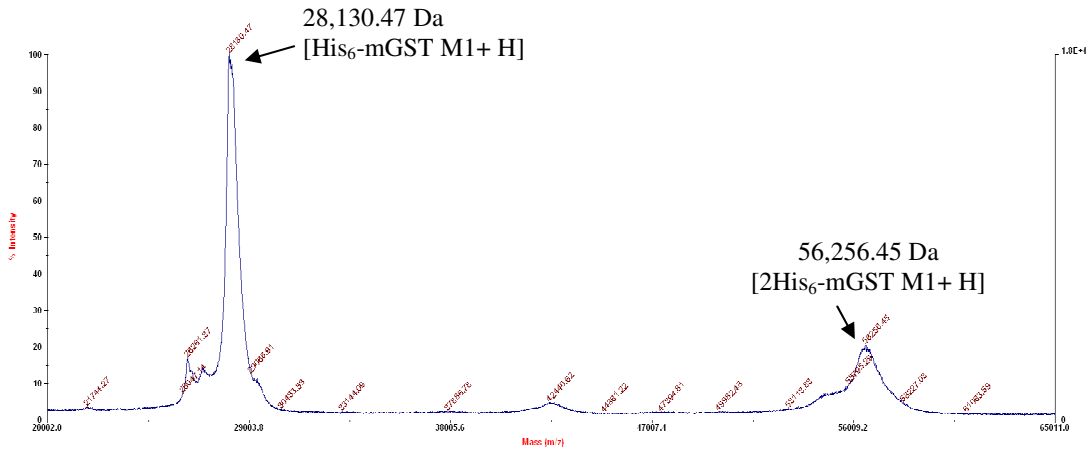


Figure 1.45. MALDI-TOF mass spectrum of the purified His_6-mGST M1-1 from 20,000 to 65,000 m/z.

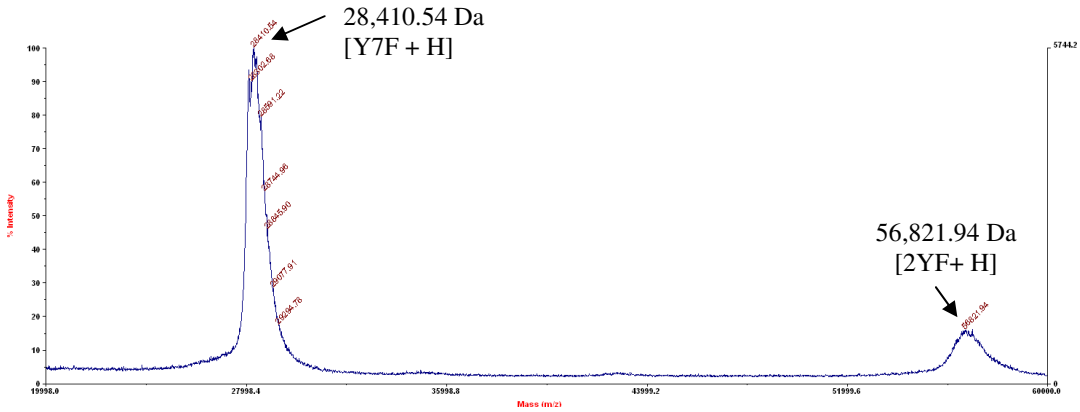


Figure 1.46. MALDI-TOF mass spectrum of the purified SjGST_Y7F from 20,000 to 65,000 m/z.

1.2.2.2. Enzymatic analyses

Using the CDNB assay specific activities of each enzyme were obtained; 105.1, 14.9, 27.2 and 137.3 μmol/min/mg for SjGST, mGST A4-4, His_6-mGST M1-1 and His_6-hGST P1-1 respectively. The data obtained for mGST A4-4 and His_6-hGST P1-1 are comparable to those found in the literature (12.2 and 130 μmol/min/mg respectively [154, 155]). Specific activities for SjGST and His_6-mGST M1-1 have not been reported in the literature. However, a value of 26 μmol/min/mg has been published for the rat GST M1-1 [74], which shares 92.6% sequence identity with mGST M1-1. Specific activity of the mutant was found to be 1.8 $\mu mol.min^{-1}.mg^{-1}$, representing only 1.7% that of the wild-type enzyme. In previous studies, García-Fuentes and co-workers generated the same mutant and found a specific activity varying between 8 and 12 $\mu mol.min^{-1}.mg^{-1}$ and a k_{cat} of 1.7% that of the wild type [156].

CDNB assays with variable concentrations of substrates (GSH: 1.0 μM to 2.0 mM; CDNB: 0.5 μM to 1.0 mM) were performed. Data were plotted with the software SigmaPlot using the Michaelis-Menten, Lineweaver-Burk and Eadie-Hofstee equations and kinetic parameters such as k_{cat} and K_M were determined. Similar results were obtained with the three models; Michaelis-Menten plots are shown in Figures 1.47. and 1.48. Kinetic data are summarised in Table 1.7.

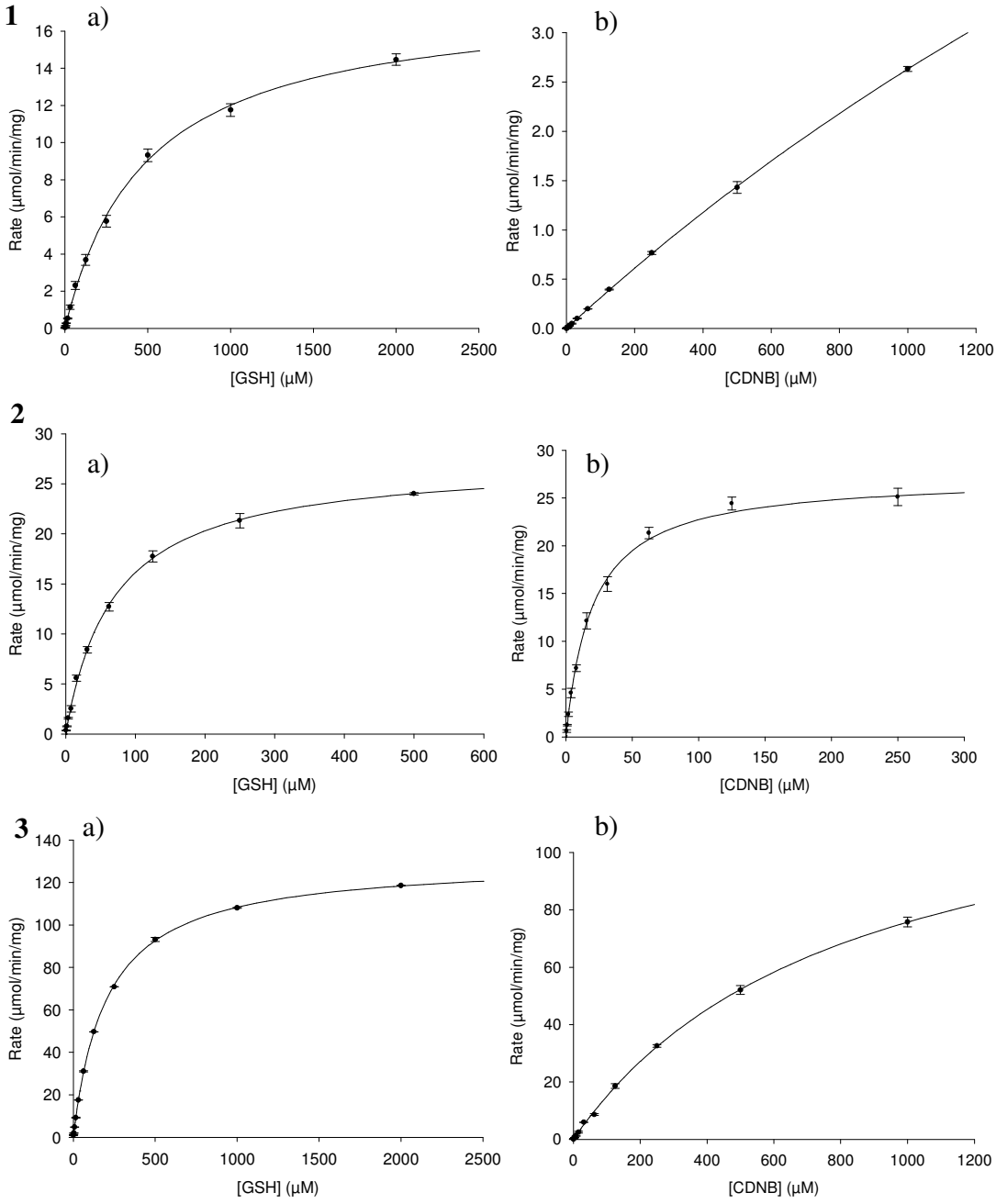


Figure 1.47. Michaelis-Menten representation of mGST A4-4 (**1**), His_6-mGST M1-1 (**2**) and His_6-hGST P1-1 (**3**) activities towards the substrates GSH (**a**) and CDNB (**b**). GSH and CDNB concentrations were varied from 1 µM to 2 mM and 0.5 µM to 1 mM, respectively.

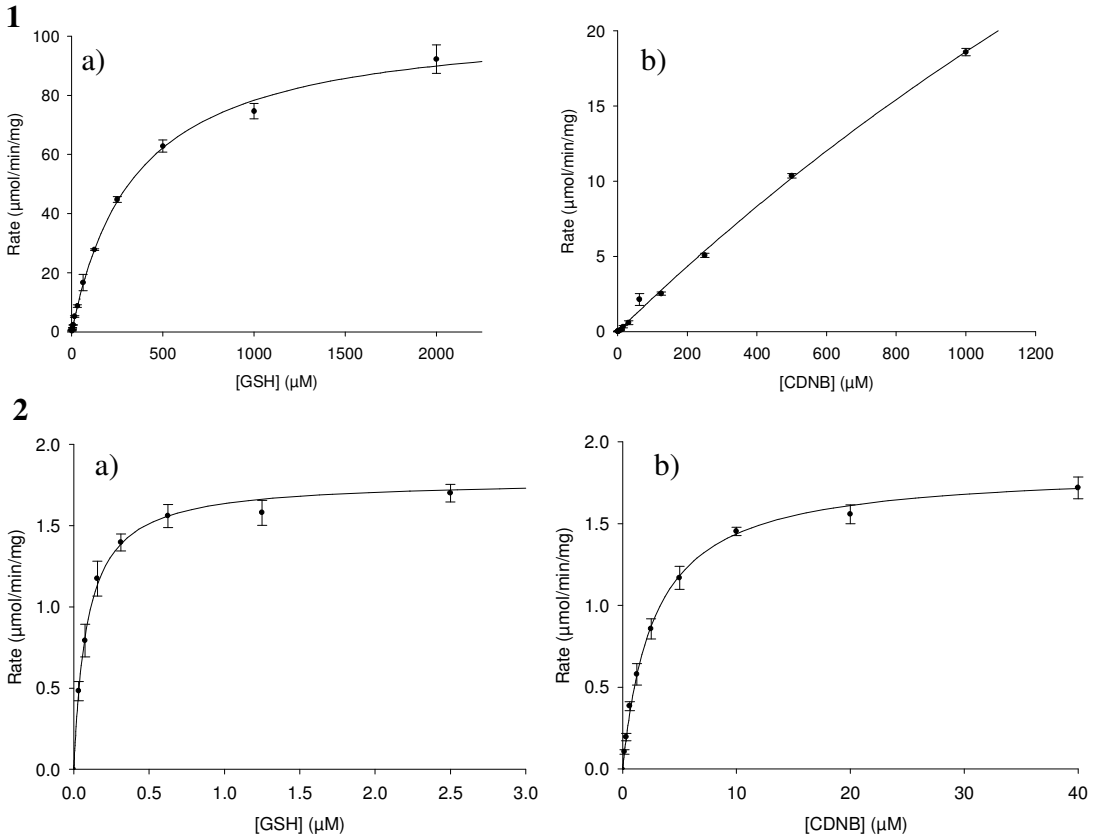


Figure 1.48. Michaelis-Menten representation of SjGST (**1**) and SjGST_Y7F (**2**) activities towards the substrates GSH (**a**) and CDNB (**b**). GSH and CDNB concentrations were varied from 1 µM to 2 mM and 0.5 µM to 1 mM respectively when assaying SjGST and from 0.03 to 10 µM and 0.4 to 40 µM respectively when assaying the mutant.

Table 1.7. Kinetic parameters for mGST A4-4, His_6-mGST M1-1, His_6-hGST P1-1, SjGST and SjGST_Y7F.

	K_M^{GSH} (mM)	K_M^{CDNB} (mM)	k_{cat}^{CDNB} (s^{-1})	k_{cat}/K_M^{CDNB} ($mM^{-1}.s^{-1}$)	Specific Activity ($\mu mol.min^{-1}.mg^{-1}$)
mGST A4-4	0.49 +/- 0.03	4.67 +/- 0.55	6.3 +/- 0.6	1.3	14.9 +/- 1.5
His_6-mGST M1-1	0.07 +/- 0.02	0.02 +/- 0.01	12.7 +/- 0.1	635	27.2 +/- 0.3
His_6-hGST P1-1	0.20 +/- 0.02	0.81 +/- 0.04	59.3 +/- 1.6	69.5	137.3 +/- 3.9
SjGST	0.35 +/- 0.03	4.65 +/- 0.92	50.0 +/- 8.3	10.8	105.1 +/- 17.6
SjGST_Y7F	0.09 +/- 0.01	2.75 +/- 0.19	0.7 +/- 0.1	0.3	1.8 +/- 0.1

The optimum conditions for the CDNB assays were obtained with a maximum CDNB concentration of 1 mM in 2.5% ethanol. CDNB precipitates at higher concentrations and a higher percentage of ethanol resulted in a decrease of the enzymes activities. The same observation was made by Jakoby and co-workers who developed the assay in 1974 [43]. Because of low solubility, the CDNB substrate concentration was less than the K_M when assaying SjGST and mGST A4-4. Therefore, the kinetic data are less accurate for these two isozymes showing higher error values.

A general trend is observed for the GST isozymes SjGST, mGST A4-4 and His_6-hGST P1-1 – the enzymes display a higher affinity to GSH than to CDNB. This is often seen with GSTs as their G-site is very well conserved and highly specific for GSH.

SjGST and mGST A4-4 have similar K_M values regarding the two substrates. Despite the fact that SjGST is most structurally similar to the mu class, it was found that its H-site bears a higher similarity to that of A4-4 [103]. These features could explain why these two enzymes have similar affinities for the same substrates.

His_6-mGST M1-1 differs from the other GSTs presented here by its relatively low affinity constants for both CDNB and GSH. By mutational experiments, Colman and co-workers have proved that the mu loop present in the rat GST M1-1 plays a crucial role in the specific recognition of CDNB [74]. Deletion of the loop substantially decreased the enzyme's affinity for the xenobiotic substrate by a 34-fold factor, giving a K_M^{CDNB} value close to the human GST P1-1's. SjGST, mGST A4-4 and His_6-hGST P1-1 all lack this structural feature.

Roberts and co-workers have shown that the extra C-terminal alpha helix (α-9) in hGST A1-1 is important for both binding and catalysis of CDNB, since its deletion induced a 26-fold increase in K_M^{CDNB} and 160-fold decrease in k_{cat}. Although A4-4 displays the same structural characteristic, it has been shown that this isoform is more specific for lipid peroxidation product 4-hydroxynonenal (4-HNE) with a specific activity 17 times higher for this substrate [157].

The relatively high affinity of His_6-hGST P1-1 for CDNB is partly due to the presence of an isoleucine in the H-site (I-105) [158], which creates a narrow active site

where the small substrate is accommodated. Mutation of this isoleucine to a valine induced a 4-fold increase in K_M^{CDNB}.

Comparable K_M values have been reported in the literature for SjGST (K_M^{GSH} = 0.43 mM, K_M^{CDNB} = 2.68; [159]), mGST A4-4 (K_M^{GSH} = 0.48 mM, K_M^{CDNB} = 4.1 mM; [160]), rGST M1-1 (K_M^{GSH} = 0.069 mM, K_M^{CDNB} = 0.019 mM; [161]) and His_6-hGST P1-1(K_M^{GSH} = 0.13 mM, K_M^{CDNB} = 0.86 mM; [154]).

The turnover number k_{cat} represents the number of substrate molecules that are converted to product per active site. In GSTs, nucleophilic aromatic substitution reactions are rate-limited by the formation of the transition state and by product release. For CDNB conjugation, the highest k_{cat} values were found for His_6-hGST P1-1 (59.3 s^{-1}) and SjGST (50.0 s^{-1}) and were 5 to 10 times lower for His_6-mGST M1-1 (12.7 s^{-1}) and mGST A4-4 (6.3 s^{-1}). Similarly, Awasthi and co-workers found a k_{cat} of 58.8 s^{-1} for hGST P1-1 [162]; and Dirr and co-workers found a k_{cat} of 18 s^{-1} for the rat GST M1-1 [163]. A slightly higher value was reported for mGST A4-4 (18.2 s^{-1}) [160]. The close proximity of the mu loop and alpha helix in mGST M1-1 and mGST A4-4 respectively, is suggested to physically restrict the rate of product release from the active sites. Furthermore, in mGST M1-1, Tyr115 located in α-4 has also been shown to play an important role in catalysis. It seems that this residue interferes with product release through hydrogen-bonding with Ser209. Mutation of Tyr115 into a Phe in the rat mGST M1-1 resulted in a 3.6-fold increase in the turnover of CDNB [164].

The catalytic efficiency of an enzyme is determined by the ratio between the rate of product release k_{cat} and the dissociation constant K_M between the enzyme and the substrate. His_6-mGST M1-1 showed the highest k_{cat}/K_M value for the GSH conjugation of CDNB when compared to the three other isoforms, indicating that, for this class of compound, the overall reaction is the most efficiently catalysed by His_6-mGST M1-1.

Mutation of the catalytically active tyrosine 7 to a phenylalanine in SjGST resulted in almost complete loss of activity toward CDNB substrate and a k_{cat} that was

1.4 % than that of the wild type. These observations show the importance of Tyr7 for the catalytic activity of this enzyme. The presence of the hydroxyl group of this tyrosine does not seem fundamental for the binding of GSH, but it would stabilize the thiolate group and give it the right orientation for the attack on the electrophilic substrate [165].

Moreover, the mutant protein showed higher affinity to GSH than the wild type. This result is in agreement with those published in 2003 by Jara-Pérez and co-workers [166] who studied the thermodynamics of GSH binding to the same mutant. They proposed that SjGST has an additional water molecule in its G-site, forming a hydrogen bond with Tyr 7 and which needs to be displaced upon GSH binding. Therefore, the balance between the number of hydrogen bonds before and after substrate binding would account for the favourable enthalpy upon binding of the mutant. This would explain the enhanced affinity for the GSH substrate to SjGST_Y7F.

With each recombinant, purified GST isoform isolated in high yield and with high catalytic activity we now proceeded to test the possibility of generating isoform-specific GST inhibitors. Previously the promiscuity of the GST H-site made rational design difficult. Protein-directed DCC provides a new method which presents the advantage to not require any structural information on the enzyme. Venughopal Bhal used a dynamic combinatorial approach to explore the active sites of the four isozymes.

1.2.3. Exploring glutathione transferase active sites

1.2.3.1. Glutathione transferase – templated dynamic combinatorial chemistry

A first dynamic combinatorial library was generated by Venughopal Bhat, based on acylhydrazone formation between an aldehyde having a CDNB scaffold and ten commercially available hydrazides. The reversible imine formation was catalysed by aniline, allowing the reaction to equilibrate rapidly at pH 6.2 [167]. Aniline functions as a nucleophilic catalyst *via* the formation of a Schiff base which then reacts with a hydrazone to yield an acyl hydrazone (Figure 1.49.). Without the aniline the equilibrium was achieved very slowly. However, in its presence it was clear that all 10 members of the DCL were observed by HPLC and mass spectrometry. Little of the initial aldehyde was observed suggesting that it had been used up in DCL formation.

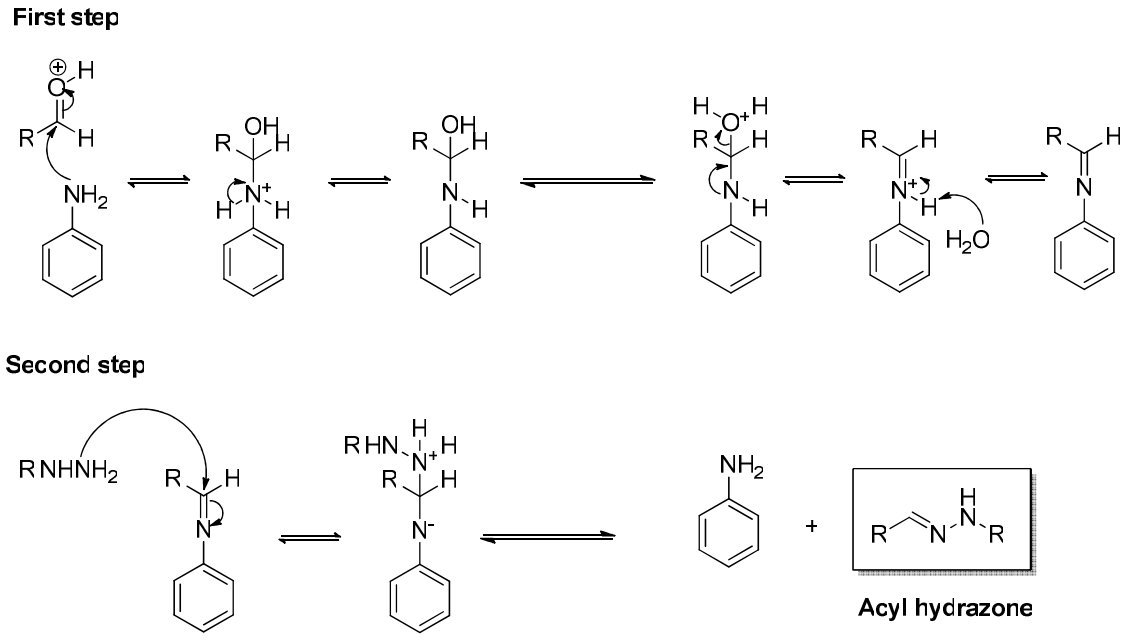


Figure 1.49. Two-step mechanism of the aniline-catalysed acyl hydrazone formation, starting from an aldehyde and a hydrazide, in acidic conditions (pH 6.2).

After verifying that SjGST retained catalytic activity in the presence of up to 20 mM of aniline, SjGST and His_6-hGST P1-1 were targeted. The acylhydrazone DCL presented in Figure 1.50. was interfaced with the two isozymes and amplification of a particular peak corresponding was observed. Both DCLs demonstrated clear amplification of hydrazone components; thiophene acylhydrazone **3g** was selected by SjGST and t-butylphenyl hydrazone **3c** by His_6-hGST P1-1 (Figure 1.51.). A control experiment was made synthesising the same DCL in the presence of bovine serum albumin (BSA) and produced no detectable amplification, indicating that GSTs isozymes were responsible for library component amplification [194]. The pH of the system was raised to 8.0 rapidly by addition of NaOH to stop the reaction and "fix" the library.

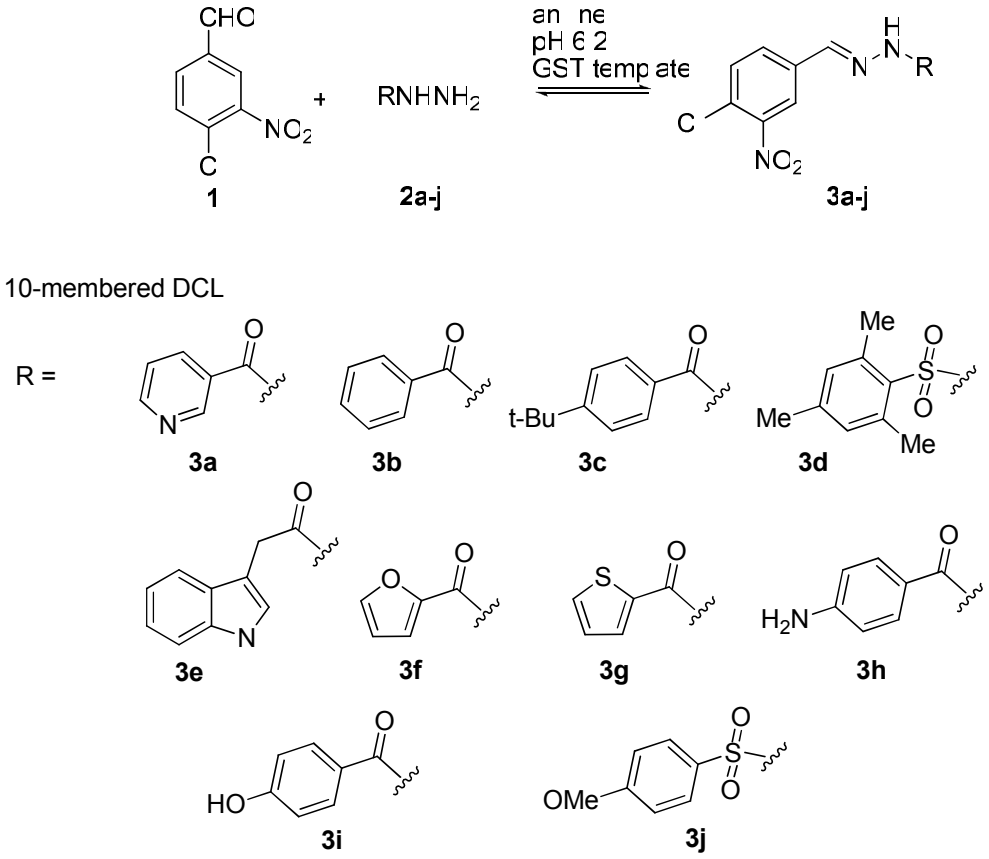


Figure 1.50. Aniline-catalysed acyl hydrazone formation. Conditions: Aldehyde (5 μM), hydrazides (20 μM) in NH_4OAc Buffer (50 mM, pH 6.2) containing 15% DMSO. Library is run in the presence of aniline (20 mM).

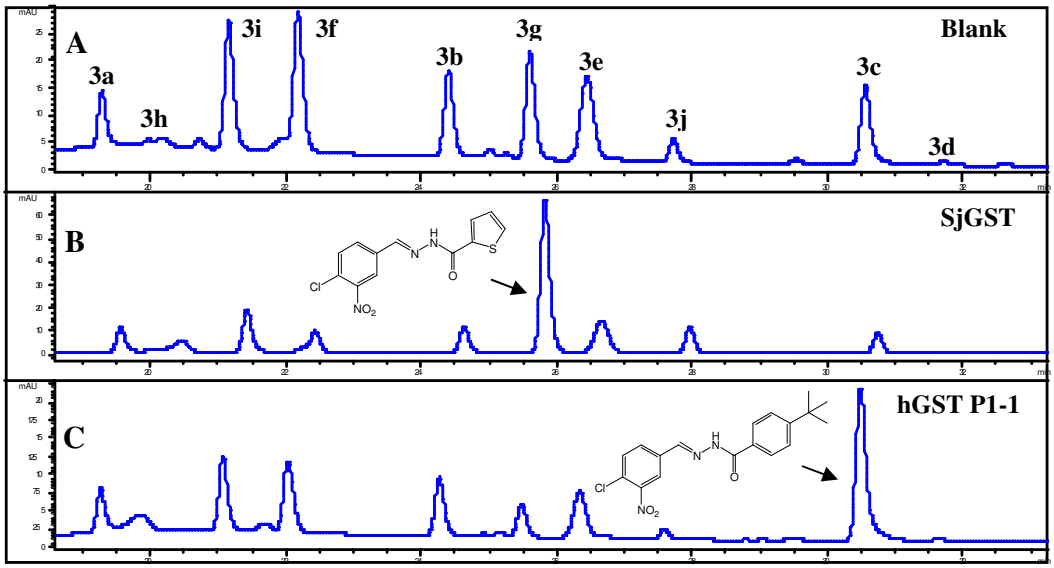


Figure 1.51. HPLC analyses of GST-templated DCLs (254 nm). **A**, DCL hydrazone composition in the absence of any target (blank). **B**, When the DCL is constituted in the presence of SjGST, the thiophene hydrazone **3g** is clearly amplified. **C**, Changing the target protein to hGSTP1-1 produces a different distribution, in which the t-butylphenyl derivative **3c** is amplified. Targeted DCL conditions: GST (1 equiv.), aldehyde (5 μM), hydrazides (20 μM) and aniline (10 mM) in NH_4OAc buffer (50 mM, pH 6.2) containing 15% DMSO for 16 h.

The amplified components were individually synthesised. I carried out some inhibition assays against both GSTs with these compounds. However, due to their poor solubility (< 10 μM), it was not possible to determine their IC_{50} values. To solve this problem and simultaneously to increase the potency of the DCL components towards the GSTs, Venughopal Bhat chemically conjugated GSH to aldehyde **1**, generating a new library composed of ten GS-conjugated hydrazones (Figure 1.52. A). We anticipated that the highly soluble GSH tripeptide motif would act as an anchor at the G-site, enabling exploration of the H-site with assorted hydrazide fragments. This approach, in which a known enzyme–substrate interaction is used for inhibitor discovery, is well exemplified in classical medicinal chemistry drug design, GST inhibition and DCC methods.

Venughopal Bhat tested the new library with the four isoforms: SjGST, His_6-hGST P1-1, His_6-mGST M1-1 and mGST A4-4. Equilibration was complete in 6 hours

in the presence of aniline at pH 6.2. As before, clear shifts in equilibrium could be observed for both SjGST and His_6-hGST P1-1 targets: in each case the same hydrazide fragment was selected as the best binder, thiophene (**5g**) for SjGST and *t*-butylphenyl (**5c**) for His_6-hGST P1-1 (Figure 1.52. B).

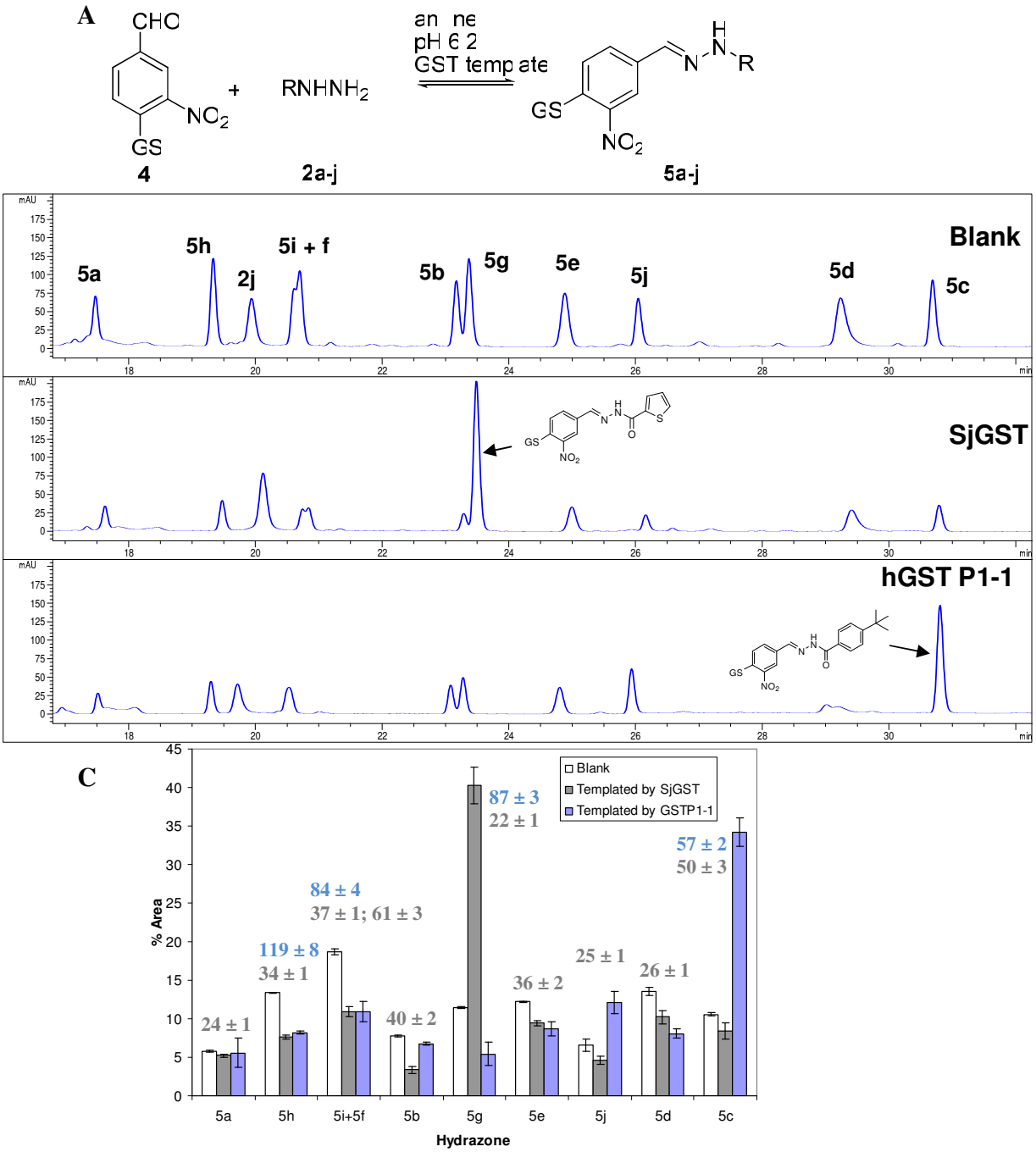


Figure 1.52. HPLC analyses of GST-templated DCLs of GSH conjugates. **A**, Acyl hydrazone DCL based on GSH-conjugated aldehyde **4** (GS-S-linked glutathione). **B**, DCL acyl hydrazone composition in the absence of target (blank), in the presence of SjGST and in the presence of hGSTP1-1. **C**, Changes in DCL component concentration for blank, SjGST and hGST P1-1 DCLs. The error bars represent the standard deviation over three experiments. IC_{50} values are indicated in grey for SjGST and in blue for hGST P1-1.

Both components were amplified to over 300% of their concentration in the blank DCL, at the expense of nearly all other competing hydrazones (Figure 1.52. C). Additionaly, the anisyl sulphonylhydrazone **5j** underwent ~100% amplification using His_6-hGST P1-1 as the only other positively selected component. The most significant reductions in equilibrium concentrations occurred for **5b**, **f** and **i** (SjGST) and **5f**, **g** and **i** (His_6-hGST P1-1).

Interestingly, the presence of either mGST A4-4 or His_6-mGST M1-1 with the GS-conjugated DCL did not induce any amplification. It is known that these two isozymes have a relatively narrow active site due to the presence of a mu loop (mGST M1-1) or an alpha helix (mGST A4-4) on the top of it. Potentially, the ten hydrazides present in the library may not be accommodated in the active sites. It is also possible that the affinities of all the library members for these enzymes are so similar that no selectivity could be detected. It would be of interest to try performing conjugation experiments with GSH to find out whether the library components can bind to the active sites of these two isoforms.

To verify the amplification results were not due to a kinetic selection by means of target-accelerated synthesis, a control experiment in which GS-conjugated DCLs were synthesised with the SjGST present from the beginning of the experiment was performed. The same equilibrium distribution was achieved, with hydrazone **5g** strongly amplified, indicating that the amplified components are the result of genuine thermodynamic selection. Further controls involved a BSA control experiment, which was negative, and DCL synthesis in the presence of a large excess of the non-selective GST inhibitor ethacrynic acid (EA) as discussed earlier. Component amplification was completely suppressed for both SjGST and His_6-hGST P1-1 DCLs, indicating that the GST active site is saturated by the EA inhibitor and cannot influence the DCL equilibrium composition.

I generated a catalytically inactive SjGST mutant to test whether a catalytically inactive enzyme would exert the same control and selectivity on DCL composition as the wild-type enzyme. Essentially zero activity was observed with this mutant in CDNB conjugation when compared with the wild-type SjGST. However, SjGST Y7F proved

equally effective in controlling DCL composition, showing a clear preference for the same thiophene derivative **5g** as was amplified by the wild-type SjGST (Figure 1.53.). This again provided evidence that the observed amplification was under thermodynamic and not kinetic control.

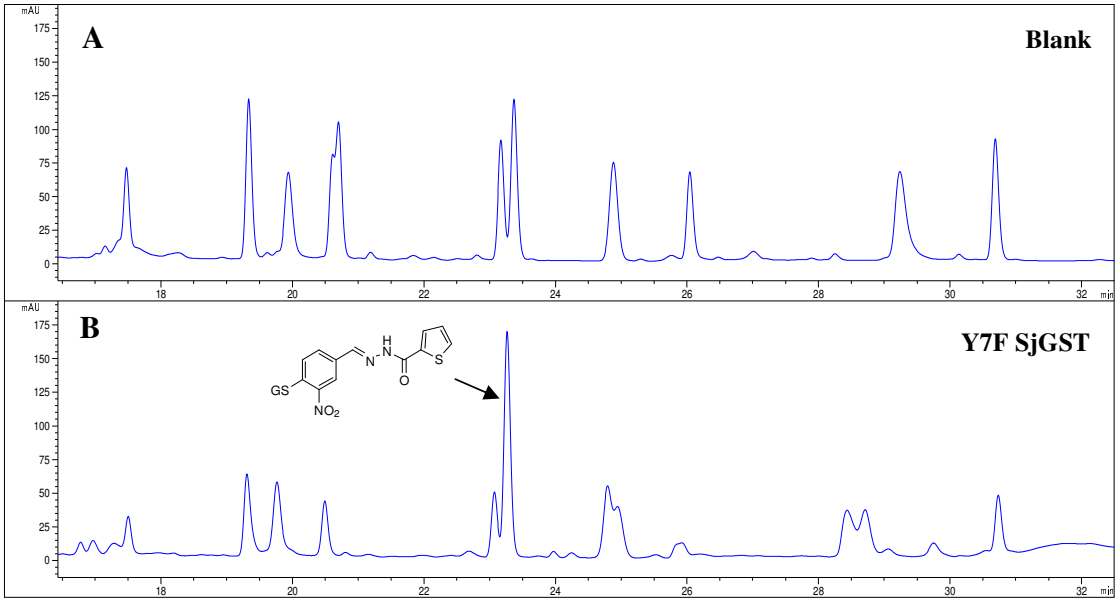


Figure 1.53. HPLC analyses of GST-templated DCLs. **A**, DCL hydrazone composition in the absence of any target (blank). **B**, When the DCL is constituted in the presence of Y7F SjGST, the thiophene hydrazone **5g** is clearly amplified.

1.2.3.2. Biological assays

I performed biological assays to establish whether the best binding compounds in the GST-directed DCLs were also the best inhibitors of the SjGST and His_6-hGST P1-1 enzymes. To fully explore the isozyme-specific amplification effects of the two DCLs, hydrazone conjugates **5a**–**5j** were separately synthesized for study.

1.2.3.2.1. Binding studies

I first confirmed that the amplified ligands **5c** and **5g** bound to SjGST and His_6-hGST P1-1 using isothermal calorimetry (ITC) and compared the thermodynamic results with those of SjGST binding to the building block **1**. ITC is a well-establish method used to measure the stoichiometry (*N*), the dissociation constant (*K*), the enthalpy (ΔH) and the entropy (ΔS) changes of a binding reaction. Typical experimental data, raw and fitted, are given in Figures 1.54. and 1.55.; the thermodynamic parameters are tabulated in Table 1.8. The interactions between the ligands and the enzymes are exothermic. Analyses of the data yielded a binding affinity (K_d) of 5.8 μM for SjGST-**5g**, 6.5 μM for SjGST-**5c**, 6.6 μM for His_6-hGSTP1-1-**5g**, 3.8 μM for His_6-hGSTP1-1-**5c**. The conjugated aldehyde **1** bound to SjGST (41 μM) with an order of magnitude less than compounds **5g** and **5c**, suggesting that the selected building blocks extend into a binding pocket and improve the potency of inhibition.

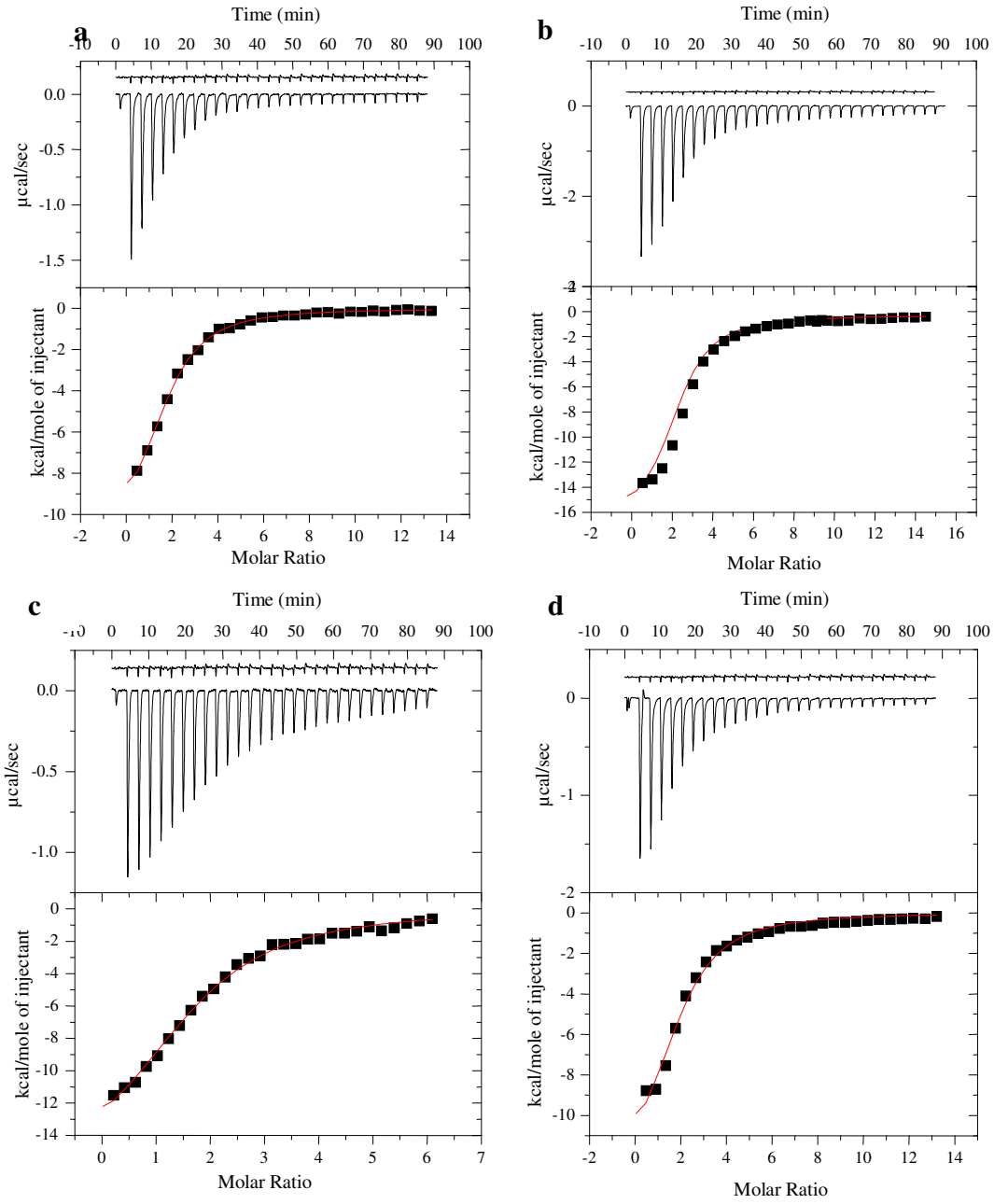


Figure 1.54. ITC of SjGST (**a** and **c**) and His_6-hGST P1-1 (**b** and **d**) in the presence and in the absence of **5c** (**c** and **d**) and **5g** (**a** and **b**). Top: Heats of injection of ~300 µM of ligand into a cell containing 10 µM of enzyme. Bottom: Data from the upper panel integrated and plotted as a function of the molar ratio of ligand after subtraction of heats generated by injection of ligand into buffer.

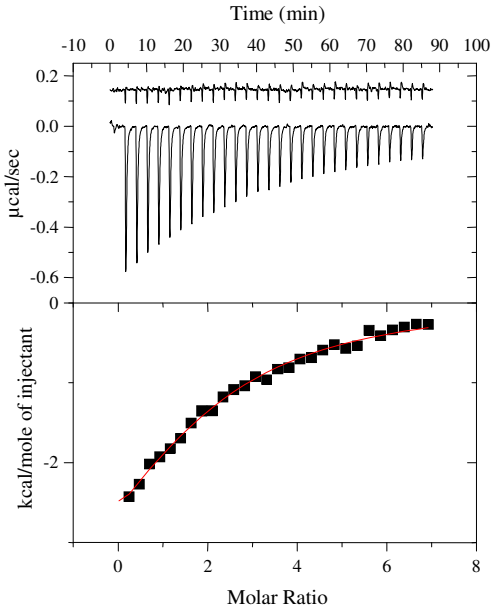


Figure 1.55. ITC of SjGST in the presence and in the absence of **1**. Top: Heats of injection of 627 µM **1** into a cell containing 20 µM SjGST. Bottom: Data from the upper panel integrated and plotted as a function of the molar ratio of **1** after subtraction of heats generated by injection of 627 µM **1** into buffer.

The stoichiometry of binding of 1.7, 1.7, 2.6 and 1.9 respectively revealed there are two ligands per dimer. The value greater than two for His_6-hGSTP1-1-**5g** may be due in part to other complexation species such as three molecules of ligand **5g** binding to a His_6-hGSTP1-1 dimer.

Table 1.8. Experimental thermodynamic data for **5c**, **5g** and **1** binding to SjGST and His_6-hGST P1-1, determined by ITC at 25 ˚C.

	SjGST-**5c**	SjGST-**5g**	hGST P1-1-**5c**	hGST P1-1-**5g**	SjGST-**1**
K_d (µM)	6.5 ± 0.3	5.8 ± 0.5	3.8 ± 0.4	6.6 ± 1.0	41 ± 3.6
N (stoichiometry)	1.71 ± 0.03	1.71 ± 0.05	2.60 ± 0.06	1.90 ± 0.1	2.02 ± 0.2
Δ*H* (kcal/mol)	-11.4 ± 0.03	-16.9 ± 0.05	-16.9 ± 0.07	-13.4 ± 0.1	-5.0 ± 0.5
Δ*S* (cal/mol)	-14.4	-33.1	-31.8	-21.4	3.2
Δ*G* (kcal/mol)	-7.1	-7.0	-7.4	-6.4	-6.0

1.2.3.2.2. Inhibition studies

I studied the inhibitory activity of the ligands towards SjGST and His_6-hGST P1-1 using the CDNB conjugation assay. The IC_{50} data are tabulated in Table 1.9.

Table 1.9. IC_{50} data (μM)

Hydrazone	sjGST	hGST P1-1
4	279 ± 23	331 ± 20
5a	24 ± 1	ND
5b	40 ± 2	ND
5c	50 ± 3	57 ± 2
5d	26 ± 1	ND
5e	36 ± 2	ND
5f	61 ± 3	ND
5g	22 ± 1	87 ± 3
5h	34 ± 1	119 ± 8
5i	37 ± 1	84 ± 4
5j	25 ± 1	ND

The IC_{50} values were slightly higher for all hydrazones against His_6-hGST P1-1 compared to SjGST (data ranging from 59 to 126 μM and 22 to 63 μM, respectively). Although all the data were found in the same micromolar range, for each isozyme, the DCC amplified hydrazone was the most active; thiophene **5g** had the lowest IC_{50} value (22 μM) among all the library members against SjGST, and t-butylphenyl **5c** had the lowest value among the four conjugates tested against His_6-hGST P1-1 (57 μM). In summary, the DCL hydrazone/GST selection process has successfully extended inhibitor structure in the GST H-site, increasing potencies by sixfold for His_6-hGST P1-1 (331 to 57 μM) and by over tenfold for SjGST (279 to 22 μM) relative to the starting anchored aldehyde **4**.

Steady-state kinetic studies on the two amplified DCL components **5c** (His_6-hGST P1-1) and **5g** (SjGST) confirmed the expected competitive inhibition profile, with both compounds binding to the GST active sites. The inhibition data for SjGST and P1-1 are plotted in Figures 1.56. and 1.57. respectively. The K_i data are presented in Table 1.10., along with the corresponding K_m and IC_{50} values. Firstly, we found that the K_i values were very close to the binding constants K_d determined previously by ITC and were well correlated to the IC_{50} data according to the Cheng-Prusoff equation: $K_i = IC_{50}/(1+[S]/K_m)$, where the K_i can be estimated as $IC_{50} = 4K_i$ for inhibitors which are competitive with both substrates [168].

Table 1.10. Ki data (μM)

	His_6-hGST P1-1		SjGST	
	CDNB	GSH	CDNB	GSH
K_m (mM)	0.81 ± 0.04	0.20 ± 0.01	4.65 ± 0.92	0.35 ± 0.02
K_i^{5g} (μM)	13.96 ± 0.78	7.19 ± 0.42	12.82 ± 0.68	5.25 ± 0.23
K_i^{5c} (μM)	10.66 ± 0.67	6.61 ± 0.42	18.58 ± 1.05	6.33 ± 0.27
IC_{50}^{5g} (μM)	87 ± 3		22 ± 1	
IC_{50}^{5c} (μM)	57 ± 2		50 ± 3	

Secondly, it was interesting to note slightly higher K_i values for both compounds when assayed against CDNB, a substrate for the H-site of the enzyme, relative to the endogenous G-site ligand GSH. The affinity of the two hydrazone conjugates towards both GST G-sites was relatively close (data ranging from 5.25 to 7.19 μM), as would be expected for two compounds sharing a common GSH-tagged nitrobenzene fragment.

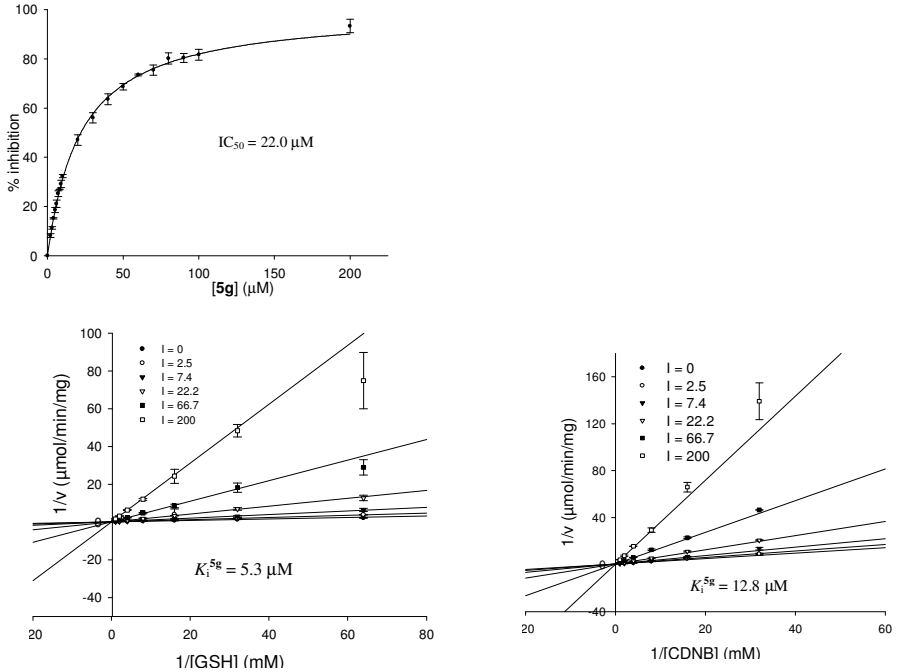


Figure 1.56. Inhibition of SjGST by **5g** using GSH and CDNB as substrates. The IC_{50} value is the concentration of inhibitor giving 50% inhibition of enzyme activity. Data are the mean ± standard deviation of triplicate experiments.

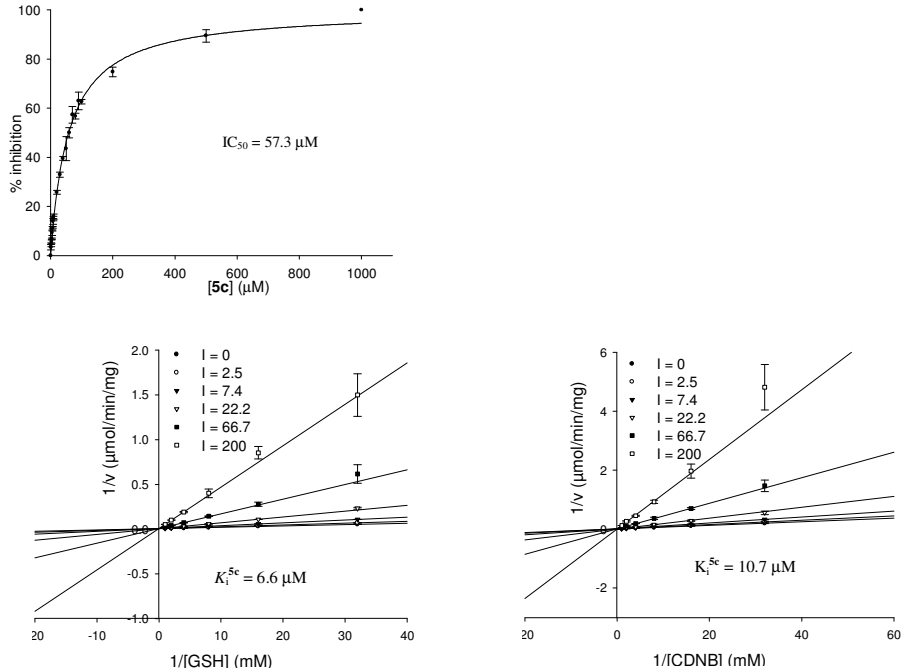


Figure 1.57. Inhibition of hGSTP1-1 by **5c** using GSH and CDNB as substrates.

1.2.3.3. Structural insights into the dynamic combinatorial library amplification mechanism

1.2.3.3.1. Docking studies

To obtain some molecular insight into the selectivity of our GST isozymes towards the two hydrazone inhibitors **5c** and **g**, I carried out a molecular modeling study with the collaboration of Dr. Ruth Brenk and Dr. Torsten Luksch (Dundee University). In a first experiment, the available GST structures were surveyed in the protein data bank (PDB) sharing more than 40% sequence identity to the target SjGST crystal structure (1M9A). 38 structures were retreived that contained a bound, GSH-based ligand. The binding sites of these structures, together with the bound ligands, were aligned, and as expected, it became evident that the GSH portions overlaid well, being bound in very similar conformations in the G-sites (Figure 1.58. a). In contrast, the conjugate parts of the various ligands showed great diversity in their conformations within the H-site, an unsurprising result given the respective functions of the G- and H-sites. Detailed analysis of the superimposed crystal structures identified the GSH conjugate of 1,2-epoxy-3-(p-nitrophenoxy)propane (EPNP) bound to cGST M1-1 (PDB code 1C72) [169] as the ligand that projected functionality into the H-site with the most similar geometry to the energy-minimized structure of hydrazone **5g** (Figure 1.58. b).

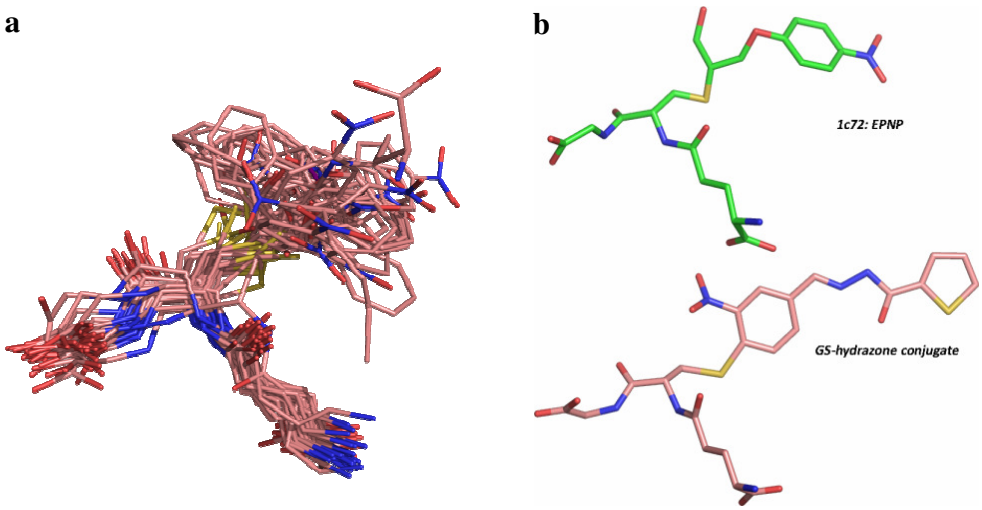


Figure 1.58. GST ligands. **a**, Superposition of a selection of GST ligands from the PDB. **b**, Conformation of the GST-bound EPNP as found in the crystal structure of cGST M1-1 (PDB code 1C72, green carbon atoms), relative to the energy-minimized structure of compound **5g** (pink carbon atoms).

Analysis of the GST–EPNP complex shows the EPNP moiety orienting towards R107 and Q165 in the H-site of the enzyme (Figure 1.59. C). The side chains of R107, F110, Q165, Q166 and F208 define the pocket that confines the EPNP moiety.

On this basis, I could generate a binding model for SjGST with thiophene hydrazone **5g** and for His_6-hGST P1-1 with t-butyl hydrazone **5c** (Figure 1.59. A and B). The interactions in the generated binding modes for SjGST in complex with **5g** and for His_6-hGST P1-1 in complex with **5c** between the glutathione moiety and the proteins are identical to those reported in previous publications [103, 108]. The hydrazone group of **5g** is predicted forms hydrogen bonds to R103 and Q204 in SjGST, and equivalent interactions are observed for **5c** in complex with His_6-hGST P1-1. Residue V161 in SjGST and I161 in His_6-hGST P1-1 make hydrophobic interactions in our models with the ligands **5c** and **5g**. As expected, the sub-pockets of both isoforms accommodating the hydrazones are rather hydrophobic, and complement the hydrophobic hydrazones amplified from the DCL. The thiophene hydrazone fits easily in the SjGST binding pocket, with only minor side-chain adjustments necessary (RMSD 0.3 Å between model

and crystal structure template), whereas the t-butylphenyl group would lead to a steric clash and would require some degree of induced fit in order to bind. Induced fit is also required to accommodate this ligand in the His_6-hGST P1-1 pocket, but in that case the binding mode could be stabilized by additional lipophilic interactions of the t-butyl group with Y103, H162 and I161. It is worth noting that Chern and colleagues have reported that mutations in that region of the H-site had a great impact on EPNP binding, with mutation of cGST M1-1 Q165 to leucine (V161 in SjGST and I161 in hGST P1-1) reducing k_{cat}^{EPNP} by 59%, although K_m showed only small changes [169]. Because the amino acids in the equivalent pocket of SjGST and His_6-hGST P1-1 are not highly conserved, these residues have such a great influence on ligand binding that it is likely that these amino-acid exchanges across the isoforms are critical in determining ligand selectivity.

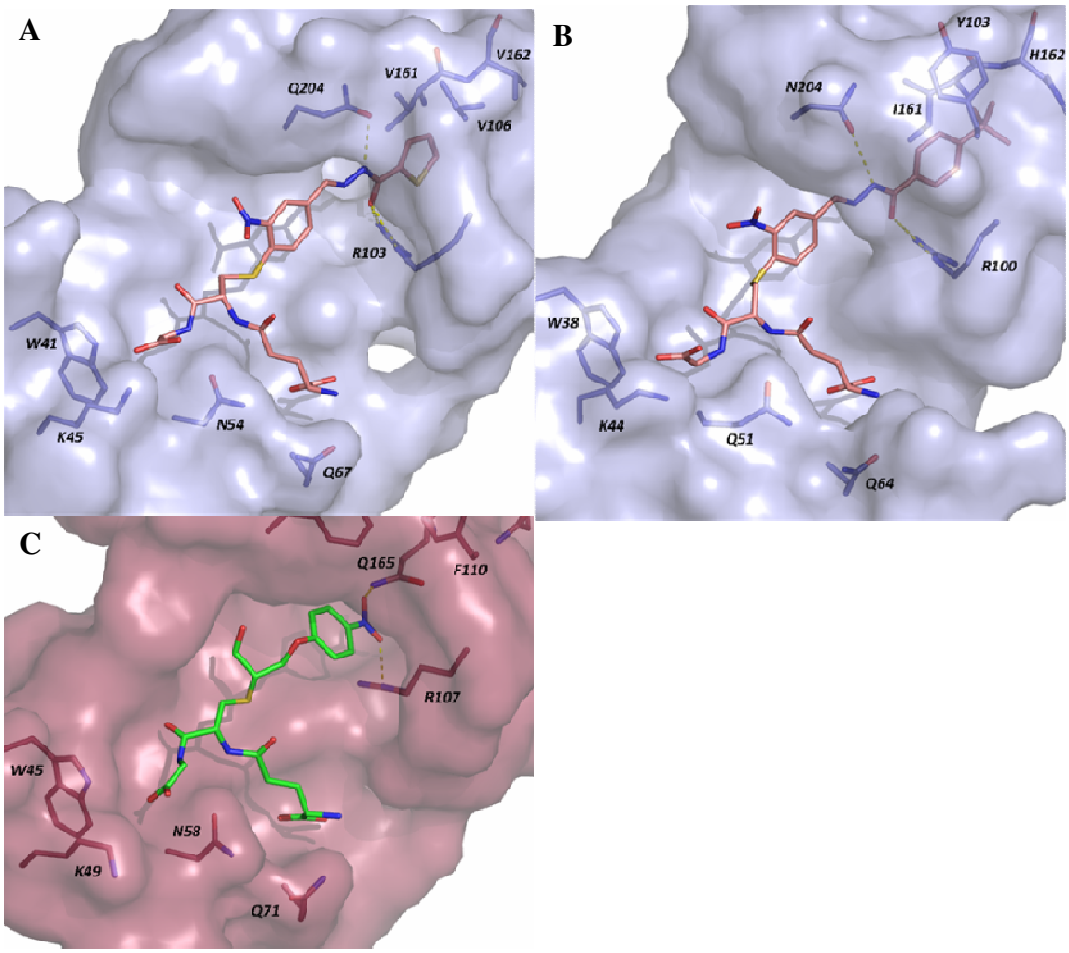


Figure 1.59. Molecular modeling of amplified DCL components with the GST active site. **A**, Model of **5g** bound to SjGST. **B**, Model of **5c** bound to His_6-hGST P1-1. The binding pocket surfaces are shown in light blue and key amino acids as blue sticks. The ligands are represented in salmon pink, with atoms coloured by type. Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines. **C**, The EPNP–cGST M1-1 crystal structure (PDB code 1C72). The binding pocket surface is shown in raspberry pink and key amino acids as red sticks. The ligand is represented in green, with atoms coloured by type. Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines.

1.2.3.3.2. Importance of the heteroatom in SjGST selection of a five-membered heterocycle

Looking back at the library members, we observed that compound **5f**, containing an oxozole ring, was fairly similar to the thiophene hydrazone **5g** selected by SjGST. Since the chemical groups of **5g** involved in the predicted interactions with SjGST are the secondary amine of the hydrazone bond, the ketone common to all the library members and C4 of the thiophene ring, we noticed that the same types of interactions could occur with **5f** and consequently it was surprising that this compound had not been amplified at all.

1.2.3.3.2.1. Generation of a 3-member dynamic combinatorial library

To study the importance of the heteroatom in the five-membered heterocycle, Venugopal Bhat generated a new library composed of the GSH conjugated aldehyde **4**, the compounds **2f**, **2g** and another acyl hydrazone containing a pyrrole ring **2k** (Figure 1.60. A). Nitrogen has an intermediate size and electronegativity between oxygen and sulfur. The new DCL was equilibrated in the absence and in the presence of SjGST. The three peaks corresponding to **5k**, **5f** and **5g** were clearly identified (Figure 1.60. B). After 16 hours with the target protein, a distinct amplification of around 50% was observed again for the same thiophene compound **5g**. Both **5k** and **5f** decreased by 55 and 50% respectively. In the acyl hydrazone **5k**, the proton bound to the nitrogen may form a hydrogen bond with the oxygen of the ketone, making this molecule more rigid and preventing interactions with R103 and Q204 in SjGST's hydrophobic pocket. Therefore binding of this compound would be less favoured among the three ligands. However, amplification of **5g** over **5k** cannot yet be explained with the current data.

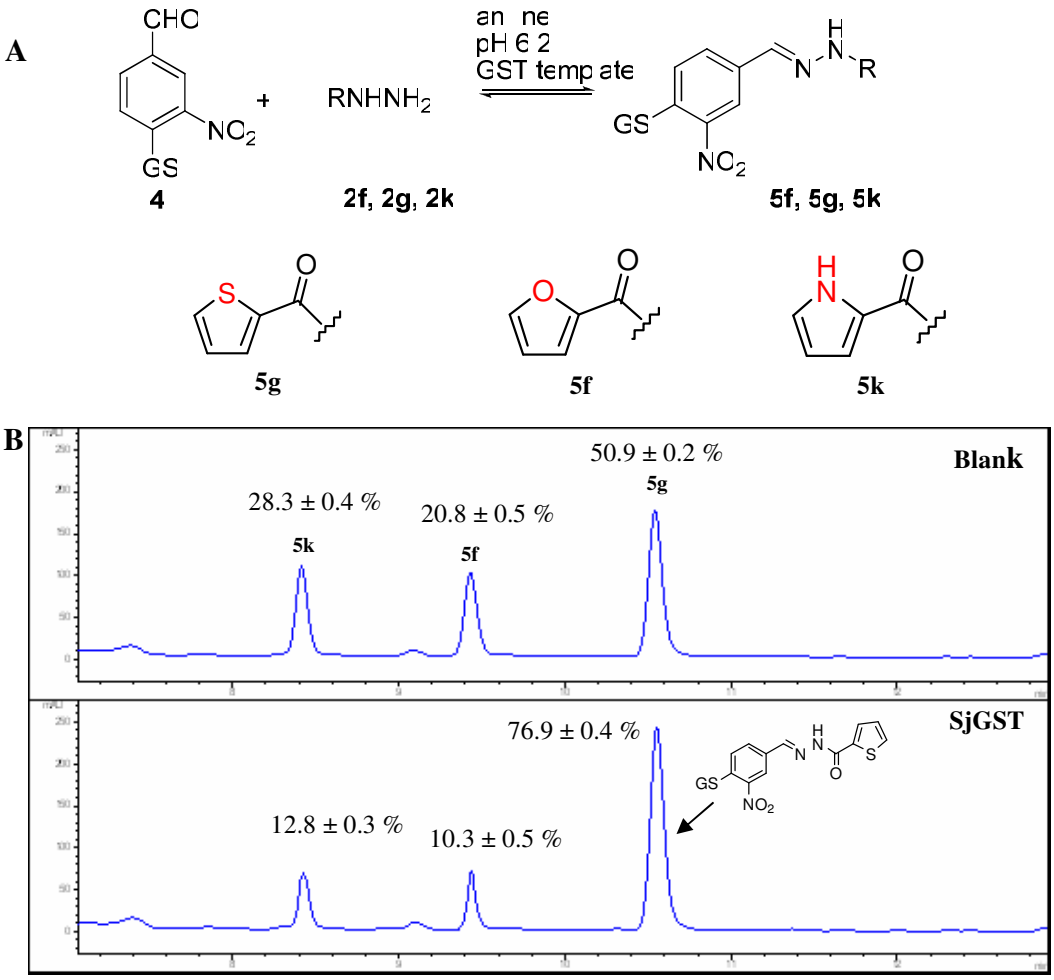


Figure 1.60. HPLC analyses of GST-templated DCL of GSH conjugates. **A**, 3-membered acyl hydrazone DCL based on GSH-conjugated aldehyde **4**. **B**, DCL acyl hydrazone composition in the absence of target (blank) and in the presence of SjGST. Percentages of peak areas are specified above each peak.

1.2.3.3.2.2. NMR studies

The structures of the five-membered heterocycle acyl hydrazones were examined by NMR. In order to simplify the NMR analyses and since we were interested in the hydrazone moiety, I analysed the non-conjugated compounds **3f** and **3g** from the first generated library. ^{1}H NMR spectra were firstly run for both compounds in DMSO at 25, 30, 35, 40 and 60°C. The spectrum of **3f** (Figure 1.61.) remained unchanged with

the increase in temperature. Whereas from 25 to 40˚C, proton peaks were too broad and coupling constants could not be determined from the **3g** spectrum (Figure 1.62. A). At 60˚C, the peaks became sharper allowing for determination of coupling constants (Figure 1.62. B), suggesting a faster rotation around the N-N bond. H7 proton was influenced most heavily by the bond delocalisation, as upon inter-conversion at higher temperature, both H7 peaks merge into an average peak hidden in the right shoulder of the peak of H3. Considering the broad peaks of H7 and H6 observed in the spectrum of **3f**, it is likely that the same dynamics occur at room temperature. The higher electronegativity of the oxygen atom may withdraw some of the electrons before the C=N-N delocalised system, allowing an easier rotation around the N-N bond. Unfortunately, ^{1}H NMR could not be run at lower temperature due the high melting point of DMSO. I suggest that **3g** may adopt a conformation more favourable to allow the thiophene ring to be accommodated within the SjGST hydrophobic pocket.

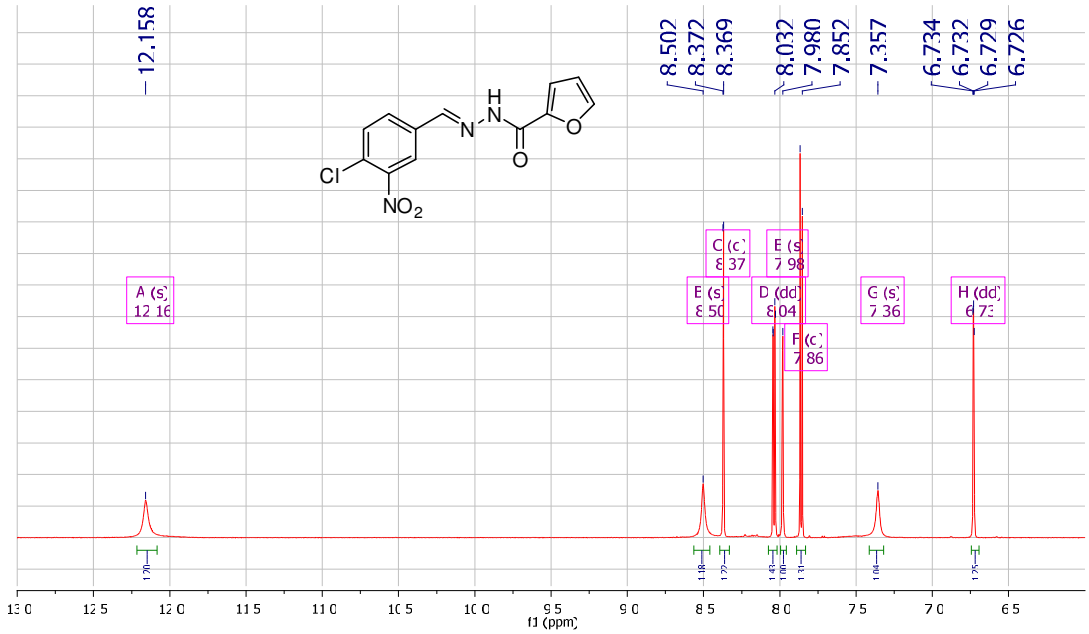


Figure 1.61. ^{1}H NMR of **3f** at room temperature

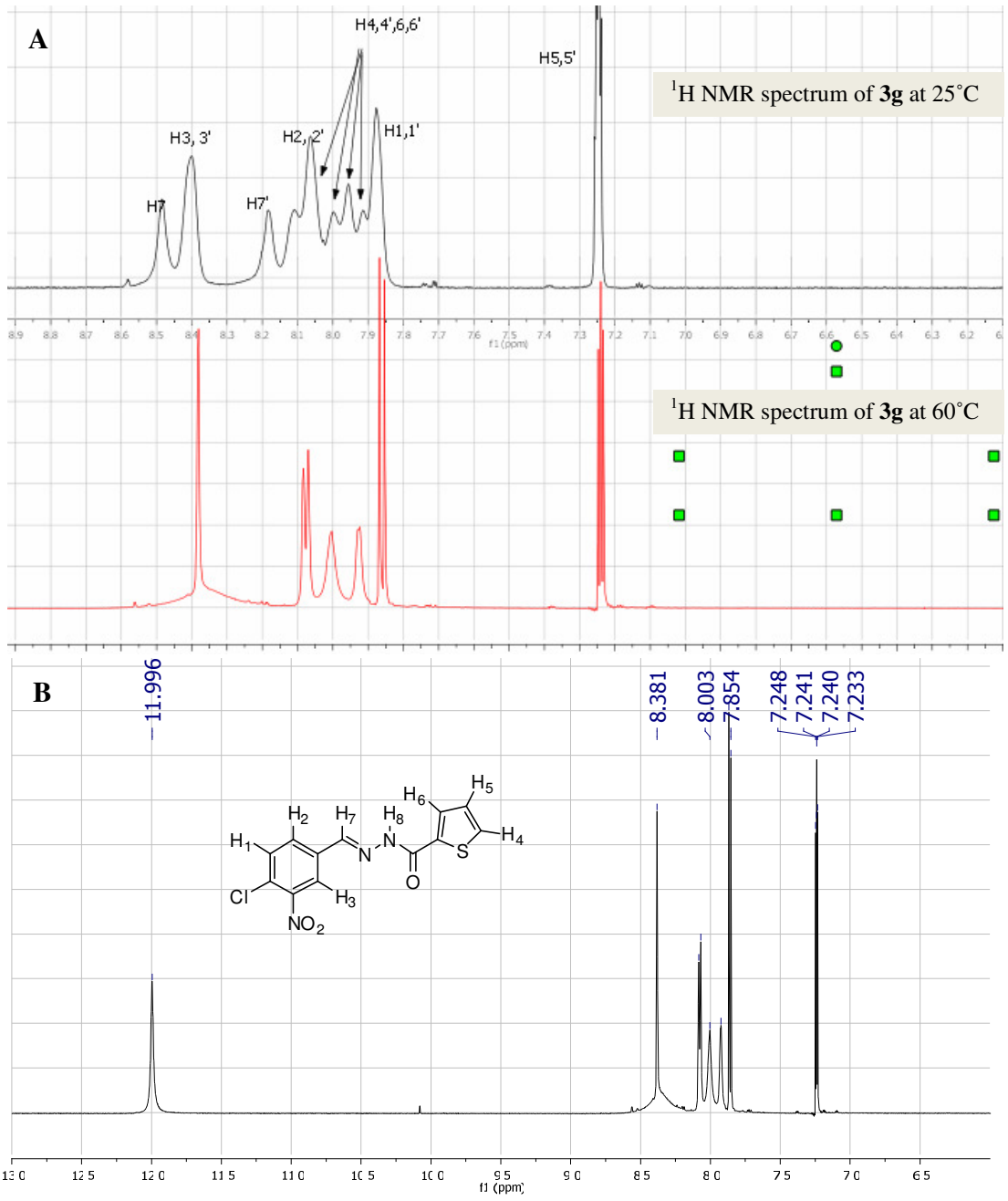


Figure 1.62. A, Superposition of **3g** ^{1}H NMR spectra at room temperature (top) and at 60˚C (bottom); **B**, Peak determination of the spectrum at 60˚C.

Currently, there are more in-depth NMR studies underway on compound **3k** to gain some insights into SjGST's selectivity.

1.3. Conclusions

GST enzymes are currently targeted in the development of new drugs against several important diseases such as cancer and schistosomiasis. In cancer therapy, they are mostly responsible for drug resistance in many cancer tissues, which often show elevated levels of theses enzymes compared to normal tissues. Many cancers also show different distribution of GST isozymes compared to those seen in healthy tissues. Therefore, elaboration of isoform-specific GST inhibitors would provide tumor-directed drugs and potentiate conventional chemotherapeutic agents.

DCC was used as a new tool to explore the promiscuous H sites of various GST isoforms. The GST enzyme proved effective as a DCL template, with two isozymes from the GST family smoothly integrating with the small molecule assemblies and strongly amplifying the best binding components. The selected hydrazones showed increased inhibitory activity of over one order of magnitude from the starting GSH-tagged nitro-benzaldehyde **4**, validating the approach in the context of protein–ligand discovery. Interestingly, a single, small DCL composed of only ten members displayed isozyme selectivity according to which variant of the GST enzyme is used as the template. To gain insight into isoform selectivity, I found that each amplified molecule could be effectively docked into its respective GST H-site, although the fine structural features of the SjGST versus hGST P1-1 H-site that discriminate between thiophene hydrazone **5g** and t-butylphenyl hydrazone **5c** are unclear at the present time. Structural determination of the various GST:GS–hydrazone conjugates will be needed for a deeper understanding of the factors that control H-site selectivity.

1.4. Chapter 1 references

1. Testa, B. and S.D. Kramer, *The biochemistry of drug metabolism--an introduction: part 1. Principles and overview.* Chem Biodivers, 2006. **3**(10): p. 1053-101.
2. Anzenbacher, P. and E. Anzenbacherova, *Cytochromes P450 and metabolism of xenobiotics.* Cell Mol Life Sci, 2001. **58**(5-6): p. 737-47.
3. Guengerich, F.P., *Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity.* Chem Res Toxicol, 2001. **14**(6): p. 611-50.
4. Guengerich, F.P., *Cytochrome p450 and chemical toxicology.* Chem Res Toxicol, 2008. **21**(1): p. 70-83.
5. Xu, C., C.Y. Li, and A.N. Kong, *Induction of phase I, II and III drug metabolism/transport by xenobiotics.* Arch Pharm Res, 2005. **28**(3): p. 249-68.
6. Schlichting, I., et al., *The catalytic pathway of cytochrome p450cam at atomic resolution.* Science, 2000. **287**(5458): p. 1615-22.
7. Jakoby, W.B., *The enzymes of detoxication.* Trans N Y Acad Sci, 1983. **41**: p. 71-5.
8. Stevens J.L., J.D.P., *The mercapturic acid pathway: biosynthesis, intermediary metabolism, and physiological disposition*, in *Glutathione. Chemical, Biochemical and Medical Aspects. Parf B*, P.R. Dolphin D., and Avramovic O., Editor. 1989, John Wiley: New York.
9. Konig, J., et al., *Conjugate export pumps of the multidrug resistance protein (MRP) family: localization, substrate specificity, and MRP2-mediated drug resistance.* Biochim Biophys Acta, 1999. **1461**(2): p. 377-94.
10. van Bladeren, P.J., *Glutathione conjugation as a bioactivation reaction.* Chem Biol Interact, 2000. **129**(1-2): p. 61-76.
11. Pickett, C.B. and A.Y. Lu, *Glutathione S-transferases: gene structure, regulation, and biological function.* Annu Rev Biochem, 1989. **58**: p. 743-64.
12. Commandeur, J.N., G.J. Stijntjes, and N.P. Vermeulen, *Enzymes and transport systems involved in the formation and disposition of glutathione S-conjugates. Role in bioactivation and detoxication mechanisms of xenobiotics.* Pharmacol Rev, 1995. **47**(2): p. 271-330.
13. Hanigan, M.H. and H.C. Pitot, *Gamma-glutamyl transpeptidase--its role in hepatocarcinogenesis.* Carcinogenesis, 1985. **6**(2): p. 165-72.
14. Whitfield, J.B., *Gamma glutamyl transferase.* Crit Rev Clin Lab Sci, 2001. **38**(4): p. 263-355.
15. Keppler, D., *Export pumps for glutathione S-conjugates.* Free Radic Biol Med, 1999. **27**(9-10): p. 985-91.
16. Yadav, S., et al., *Linking stress-signaling, glutathione metabolism, signaling pathways and xenobiotic transporters.* Cancer Metastasis Rev, 2007. **26**(1): p. 59-69.

17. McIntyre, T. and N.P. Curthoys, *Renal catabolism of glutathione. Characterization of a particulate rat renal dipeptidase that catalyzes the hydrolysis of cysteinylglycine.* J Biol Chem, 1982. **257**(20): p. 11915-21.
18. Duffel, M.W. and W.B. Jakoby, *Cysteine S-conjugate N-acetyltransferase from rat kidney microsomes.* Mol Pharmacol, 1982. **21**(2): p. 444-8.
19. Hopkins, F.G., *On an Autoxidisable Constituent of the Cell.* Biochem J, 1921. **15**(2): p. 286-305.
20. Nicolet, B.H., *The Structure of Glutathione.* Science, 1930. **71**(1849): p. 589-90.
21. Dalle-Donne, I., et al., *Protein S-glutathionylation: a regulatory device from bacteria to humans.* Trends Biochem Sci, 2009. **34**(2): p. 85-96.
22. Newton, G.L., N. Buchmeier, and R.C. Fahey, *Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria.* Microbiol Mol Biol Rev, 2008. **72**(3): p. 471-94.
23. Newton, G.L., et al., *Bacillithiol is an antioxidant thiol produced in Bacilli.* Nat Chem Biol, 2009. **5**(9): p. 625-7.
24. Meister, A., *Glutathione metabolism and its selective modification.* J Biol Chem, 1988. **263**(33): p. 17205-8.
25. Ketterer, B., B. Coles, and D.J. Meyer, *The role of glutathione in detoxication.* Environ Health Perspect, 1983. **49**: p. 59-69.
26. Lu, S.C., *Regulation of glutathione synthesis.* Mol Aspects Med, 2009. **30**(1-2): p. 42-59.
27. Sies, H., *Glutathione and its role in cellular functions.* Free Radic Biol Med, 1999. **27**(9-10): p. 916-21.
28. Meister, A. and M.E. Anderson, *Glutathione.* Annu Rev Biochem, 1983. **52**: p. 711-60.
29. Griffith, O.W., *Biologic and pharmacologic regulation of mammalian glutathione synthesis.* Free Radic Biol Med, 1999. **27**(9-10): p. 922-35.
30. Forman, H.J., H. Zhang, and A. Rinna, *Glutathione: overview of its protective roles, measurement, and biosynthesis.* Mol Aspects Med, 2009. **30**(1-2): p. 1-12.
31. Franklin, C.C., et al., *Structure, function, and post-translational regulation of the catalytic and modifier subunits of glutamate cysteine ligase.* Mol Aspects Med, 2009. **30**(1-2): p. 86-98.
32. Hibi, T., et al., *Crystal structure of gamma-glutamylcysteine synthetase: insights into the mechanism of catalysis by a key enzyme for glutathione homeostasis.* Proc Natl Acad Sci U S A, 2004. **101**(42): p. 15052-7.
33. Matsuda, K., et al., *Crystal structure of glutathione synthetase at optimal pH: domain architecture and structural similarity with other proteins.* Protein Eng, 1996. **9**(12): p. 1083-92.
34. Hwang, C., A.J. Sinskey, and H.F. Lodish, *Oxidized redox state of glutathione in the endoplasmic reticulum.* Science, 1992. **257**(5076): p. 1496-502.
35. Meredith, M.J. and D.J. Reed, *Status of the mitochondrial pool of glutathione in the isolated hepatocyte.* J Biol Chem, 1982. **257**(7): p. 3747-53.

36. Perricone, C., C. De Carolis, and R. Perricone, *Glutathione: a key player in autoimmunity.* Autoimmun Rev, 2009. **8**(8): p. 697-701.
37. Bindoli, A., J.M. Fukuto, and H.J. Forman, *Thiol chemistry in peroxidase catalysis and redox signaling.* Antioxid Redox Signal, 2008. **10**(9): p. 1549-64.
38. Akerboom, T.P., M. Bilzer, and H. Sies, *The relationship of biliary glutathione disulfide efflux and intracellular glutathione disulfide content in perfused rat liver.* J Biol Chem, 1982. **257**(8): p. 4248-52.
39. Meyer, A.J., *The integration of glutathione homeostasis and redox signaling.* J Plant Physiol, 2008. **165**(13): p. 1390-403.
40. Meister, A., *Glutathione, ascorbate, and cellular protection.* Cancer Res, 1994. **54**(7 Suppl): p. 1969s-1975s.
41. Kirlin, W.G., et al., *Glutathione redox potential in response to differentiation and enzyme inducers.* Free Radic Biol Med, 1999. **27**(11-12): p. 1208-18.
42. Dalle-Donne, I., et al., *S-glutathionylation in protein redox regulation.* Free Radic Biol Med, 2007. **43**(6): p. 883-98.
43. Habig, W.H., M.J. Pabst, and W.B. Jakoby, *Glutathione S-transferases. The first enzymatic step in mercapturic acid formation.* J Biol Chem, 1974. **249**(22): p. 7130-9.
44. Mahajan, S. and W.M. Atkins, *The chemistry and biology of inhibitors and pro-drugs targeted to glutathione S-transferases.* Cell Mol Life Sci, 2005. **62**(11): p. 1221-33.
45. Booth, J., E. Boyland, and P. Sims, *An enzyme from rat liver catalysing conjugations with glutathione.* Biochem J, 1961. **79**(3): p. 516-24.
46. Armstrong, R.N., *Structure, catalytic mechanism, and evolution of the glutathione transferases.* Chem Res Toxicol, 1997. **10**(1): p. 2-18.
47. Armstrong, R.N., *Mechanistic imperatives for the evolution of glutathione transferases.* Curr Opin Chem Biol, 1998. **2**(5): p. 618-23.
48. Mannervik, B. and U.H. Danielson, *Glutathione transferases--structure and catalytic activity.* CRC Crit Rev Biochem, 1988. **23**(3): p. 283-337.
49. Hayes, J.D., J.U. Flanagan, and I.R. Jowsey, *Glutathione transferases.* Annu Rev Pharmacol Toxicol, 2005. **45**: p. 51-88.
50. Oakley, A.J., *Glutathione transferases: new functions.* Curr Opin Struct Biol, 2005. **15**(6): p. 716-23.
51. Dixon, D.P., M. Skipsey, and R. Edwards, *Roles for glutathione transferases in plant secondary metabolism.* Phytochemistry.
52. Fortin, P.D., et al., *A glutathione S-transferase catalyzes the dehalogenation of inhibitory metabolites of polychlorinated biphenyls.* J Bacteriol, 2006. **188**(12): p. 4424-30.
53. Sheehan, D., et al., *Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily.* Biochem J, 2001. **360**(Pt 1): p. 1-16.
54. Martin, J.L., *Thioredoxin--a fold for all reasons.* Structure, 1995. **3**(3): p. 245-50.
55. Wedemeyer, W.J., E. Welker, and H.A. Scheraga, *Proline cis-trans isomerization and protein folding.* Biochemistry, 2002. **41**(50): p. 14637-44.

56. Dirr, H., P. Reinemer, and R. Huber, *X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function.* Eur J Biochem, 1994. **220**(3): p. 645-61.
57. Patskovsky, Y., et al., *Transition state model and mechanism of nucleophilic aromatic substitution reactions catalyzed by human glutathione S-transferase M1a-1a.* Biochemistry, 2006. **45**(12): p. 3852-62.
58. Manoharan, T.H., et al., *Structural studies on human glutathione S-transferase pi. Substitution mutations to determine amino acids necessary for binding glutathione.* J Biol Chem, 1992. **267**(26): p. 18940-5.
59. Stenberg, G., P.G. Board, and B. Mannervik, *Mutation of an evolutionarily conserved tyrosine residue in the active site of a human class Alpha glutathione transferase.* FEBS Lett, 1991. **293**(1-2): p. 153-5.
60. Dourado, D.F., et al., *Glutathione transferase: new model for glutathione activation.* Chemistry, 2008. **14**(31): p. 9591-8.
61. Board, P.G., et al., *Evidence for an essential serine residue in the active site of the Theta class glutathione transferases.* Biochem J, 1995. **311 (Pt 1)**: p. 247-50.
62. Flanagan, J.U., et al., *Mutagenic analysis of conserved arginine residues in and around the novel sulfate binding pocket of the human Theta class glutathione transferase T2-2.* Protein Sci, 1999. **8**(10): p. 2205-12.
63. Nishida, M., et al., *Three-dimensional structure of Escherichia coli glutathione S-transferase complexed with glutathione sulfonate: catalytic roles of Cys10 and His106.* J Mol Biol, 1998. **281**(1): p. 135-47.
64. Tocheva, E.I., et al., *Structures of ternary complexes of BphK, a bacterial glutathione S-transferase that reductively dechlorinates polychlorinated biphenyl metabolites.* J Biol Chem, 2006. **281**(41): p. 30933-40.
65. Ketley, J.N., W.H. Habig, and W.B. Jakoby, *Binding of nonsubstrate ligands to the glutathione S-transferases.* J Biol Chem, 1975. **250**(22): p. 8670-3.
66. Habig, W.H., et al., *The identity of glutathione S-transferase B with ligandin, a major binding protein of liver.* Proc Natl Acad Sci U S A, 1974. **71**(10): p. 3879-82.
67. Oakley, A.J., et al., *The ligandin (non-substrate) binding site of human Pi class glutathione transferase is located in the electrophile binding site (H-site).* J Mol Biol, 1999. **291**(4): p. 913-26.
68. McTigue, M.A., D.R. Williams, and J.A. Tainer, *Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica and its complex with the leading antischistosomal drug praziquantel.* J Mol Biol, 1995. **246**(1): p. 21-7.
69. Ji, X., et al., *Structure and function of the xenobiotic substrate-binding site and location of a potential non-substrate-binding site in a class pi glutathione S-transferase.* Biochemistry, 1997. **36**(32): p. 9690-702.
70. Townsend, D.M. and K.D. Tew, *The role of glutathione-S-transferase in anti-cancer drug resistance.* Oncogene, 2003. **22**(47): p. 7369-75.
71. McLellan, L.I. and C.R. Wolf, *Glutathione and glutathione-dependent enzymes in cancer drug resistance.* Drug Resist Updat, 1999. **2**(3): p. 153-164.

72. Simon, S.M. and M. Schindler, *Cell biological mechanisms of multidrug resistance in tumors.* Proc Natl Acad Sci U S A, 1994. **91**(9): p. 3497-504.
73. Tew, K.D., S. Dutta, and M. Schultz, *Inhibitors of glutathione S-transferases as therapeutic agents.* Adv Drug Deliv Rev, 1997. **26**(2-3): p. 91-104.
74. Hearne, J.L. and R.F. Colman, *Contribution of the mu loop to the structure and function of rat glutathione transferase M1-1.* Protein Sci, 2006. **15**(6): p. 1277-89.
75. Gustafsson, A., et al., *The C-terminal region of human glutathione transferase A1-1 affects the rate of glutathione binding and the ionization of the active-site Tyr9.* Biochemistry, 1999. **38**(49): p. 16268-75.
76. Shokeer, A., A.K. Larsson, and B. Mannervik, *Residue 234 in glutathione transferase T1-1 plays a pivotal role in the catalytic activity and the selectivity against alternative substrates.* Biochem J, 2005. **388**(Pt 1): p. 387-92.
77. Rossjohn, J., et al., *Human theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site.* Structure, 1998. **6**(3): p. 309-22.
78. Hussey, A.J. and J.D. Hayes, *Characterization of a human class-Theta glutathione S-transferase with activity towards 1-menaphthyl sulphate.* Biochem J, 1992. **286 (Pt 3)**: p. 929-35.
79. Gillham, B., *The reaction of aralkyl sulphate esters with glutathione catalysed by rat liver preparations.* Biochem J, 1971. **121**(4): p. 667-72.
80. Board, P.G., et al., *Identification, characterization, and crystal structure of the Omega class glutathione transferases.* J Biol Chem, 2000. **275**(32): p. 24798-806.
81. Dulhunty, A., et al., *The glutathione transferase structural family includes a nuclear chloride channel and a ryanodine receptor calcium release channel modulator.* J Biol Chem, 2001. **276**(5): p. 3319-23.
82. Ji, X., et al., *Location of a potential transport binding site in a sigma class glutathione transferase by x-ray crystallography.* Proc Natl Acad Sci U S A, 1996. **93**(16): p. 8208-13.
83. Thomson, A.M., D.J. Meyer, and J.D. Hayes, *Sequence, catalytic properties and expression of chicken glutathione-dependent prostaglandin D2 synthase, a novel class Sigma glutathione S-transferase.* Biochem J, 1998. **333 (Pt 2)**: p. 317-25.
84. Meyer, D.J., et al., *Purification and characterization of prostaglandin-H E-isomerase, a sigma-class glutathione S-transferase, from Ascaridia galli.* Biochem J, 1996. **313 (Pt 1)**: p. 223-7.
85. Weber, J.E., et al., *Identification and characterisation of new inhibitors for the human hematopoietic prostaglandin D2 synthase.* Eur J Med Chem. **45**(2): p. 447-54.
86. Polekhina, G., et al., *Crystal structure of maleylacetoacetate isomerase/glutathione transferase zeta reveals the molecular basis for its remarkable catalytic promiscuity.* Biochemistry, 2001. **40**(6): p. 1567-76.
87. Board, P.G., et al., *Zeta, a novel class of glutathione transferases in a range of species from plants to humans.* Biochem J, 1997. **328 (Pt 3)**: p. 929-35.

88. Tong, Z., P.G. Board, and M.W. Anders, *Glutathione transferase zeta catalyses the oxygenation of the carcinogen dichloroacetic acid to glyoxylic acid.* Biochem J, 1998. **331 (Pt 2)**: p. 371-4.
89. Gulick, A.M. and W.E. Fahl, *Mammalian glutathione S-transferase: regulation of an enzyme system to achieve chemotherapeutic efficacy.* Pharmacol Ther, 1995. **66**(2): p. 237-57.
90. Meyer, D.J., et al., *Chlorambucil-monoglutathionyl conjugate is sequestered by human alpha class glutathione S-transferases.* Br J Cancer, 1992. **66**(3): p. 433-8.
91. Riddick, D.S., et al., *Cancer chemotherapy and drug metabolism.* Drug Metab Dispos, 2005. **33**(8): p. 1083-96.
92. Stavrovskaya, A.A., *Cellular mechanisms of multidrug resistance of tumor cells.* Biochemistry (Mosc), 2000. **65**(1): p. 95-106.
93. Lewis, A.D., et al., *Amplification and increased expression of alpha class glutathione S-transferase-encoding genes associated with resistance to nitrogen mustards.* Proc Natl Acad Sci U S A, 1988. **85**(22): p. 8511-5.
94. Hu, W., et al., *Schistosome transcriptomes: new insights into the parasite and schistosomiasis.* Trends Mol Med, 2004. **10**(5): p. 217-25.
95. McManus, D.P. and A. Loukas, *Current status of vaccines for schistosomiasis.* Clin Microbiol Rev, 2008. **21**(1): p. 225-42.
96. Organisation, W.H. *Schistosomiasis*. 2010 [cited 2010 18-03-2010]; Available from: http://www.who.int/schistosomiasis/en/.
97. Caffrey, C.R., *Chemotherapy of schistosomiasis: present and future.* Curr Opin Chem Biol, 2007. **11**(4): p. 433-9.
98. Brindley, P.J., *The molecular biology of schistosomes.* Trends Parasitol, 2005. **21**(11): p. 533-6.
99. Kaplan, W., et al., *Conformational stability of pGEX-expressed Schistosoma japonicum glutathione S-transferase: a detoxification enzyme and fusion-protein affinity tag.* Protein Sci, 1997. **6**(2): p. 399-406.
100. Brophy, P.M. and D.I. Pritchard, *Parasitic helminth glutathione S-transferases: an update on their potential as targets for immuno- and chemotherapy.* Exp Parasitol, 1994. **79**(1): p. 89-96.
101. Torres-Rivera, A. and A. Landa, *Glutathione transferases from parasites: a biochemical view.* Acta Trop, 2008. **105**(2): p. 99-112.
102. Rufer, A.C., et al., *X-ray structure of glutathione S-transferase from Schistosoma japonicum in a new crystal form reveals flexibility of the substrate-binding site.* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2005. **61**(Pt 3): p. 263-5.
103. Cardoso, R.M., et al., *Characterization of the electrophile binding site and substrate binding mode of the 26-kDa glutathione S-transferase from Schistosoma japonicum.* Proteins, 2003. **51**(1): p. 137-46.
104. Lim, K., et al., *Three-dimensional structure of Schistosoma japonicum glutathione S-transferase fused with a six-amino acid conserved neutralizing epitope of gp41 from HIV.* Protein Sci, 1994. **3**(12): p. 2233-44.

105. Milhon, J.L., et al., *Schistosoma japonicum GSH S-transferase Sj26 is not the molecular target of praziquantel action.* Exp Parasitol, 1997. **87**(3): p. 268-74.
106. Bosher, S.K., *The nature of the ototoxic actions of ethacrynic acid upon the mammalian endolymph system. I. Functional aspects.* Acta Otolaryngol, 1980. **89**(5-6): p. 407-18.
107. Zhao, G., et al., *Synthesis and structure-activity relationship of ethacrynic acid analogues on glutathione-s-transferase P1-1 activity inhibition.* Bioorg Med Chem, 2005. **13**(12): p. 4056-62.
108. Oakley, A.J., et al., *The three-dimensional structure of the human Pi class glutathione transferase P1-1 in complex with the inhibitor ethacrynic acid and its glutathione conjugate.* Biochemistry, 1997. **36**(3): p. 576-85.
109. Oakley, A.J., et al., *The glutathione conjugate of ethacrynic acid can bind to human pi class glutathione transferase P1-1 in two different modes.* FEBS Lett, 1997. **419**(1): p. 32-6.
110. Tew, K.D., A.M. Bomber, and S.J. Hoffman, *Ethacrynic acid and piriprost as enhancers of cytotoxicity in drug resistant and sensitive cell lines.* Cancer Res, 1988. **48**(13): p. 3622-5.
111. Hayes, J.D., *Selective elution of rodent glutathione S-transferases and glyoxalase I from the S-hexyglutathione-Sepharose affinity matrix.* Biochem J, 1988. **255**(3): p. 913-22.
112. Lyon, R.P., J.J. Hill, and W.M. Atkins, *Novel class of bivalent glutathione S-transferase inhibitors.* Biochemistry, 2003. **42**(35): p. 10418-28.
113. Maeda, D.Y., et al., *Bivalent inhibitors of glutathione S-transferase: the effect of spacer length on isozyme selectivity.* Bioorg Med Chem Lett, 2006. **16**(14): p. 3780-3.
114. Mahajan, S.S., et al., *Optimization of bivalent glutathione S-transferase inhibitors by combinatorial linker design.* J Am Chem Soc, 2006. **128**(26): p. 8615-25.
115. Morgan, A.S., et al., *Tumor efficacy and bone marrow-sparing properties of TER286, a cytotoxin activated by glutathione S-transferase.* Cancer Res, 1998. **58**(12): p. 2568-75.
116. Satyam, A., et al., *Design, synthesis, and evaluation of latent alkylating agents activated by glutathione S-transferase.* J Med Chem, 1996. **39**(8): p. 1736-47.
117. Findlay, V.J., et al., *Tumor cell responses to a novel glutathione S-transferase-activated nitric oxide-releasing prodrug.* Mol Pharmacol, 2004. **65**(5): p. 1070-9.
118. Eklund, B.I., et al., *Human glutathione transferases catalyzing the bioactivation of anticancer thiopurine prodrugs.* Biochem Pharmacol, 2007. **73**(11): p. 1829-41.
119. Rosario, L.A., et al., *Cellular response to a glutathione S-transferase P1-1 activated prodrug.* Mol Pharmacol, 2000. **58**(1): p. 167-74.
120. Otto, S., R.L. Furlan, and J.K. Sanders, *Dynamic combinatorial chemistry.* Drug Discov Today, 2002. **7**(2): p. 117-25.
121. Ramstrom, O. and J.M. Lehn, *Drug discovery by dynamic combinatorial libraries.* Nat Rev Drug Discov, 2002. **1**(1): p. 26-36.

122. Bhat V.T., G.M.F., *Protein-Directed Dynamic Combinatorial Chemistry*, in *Dynamic Combinatorial Chemistry*, B.L. Miller, Editor. 2009, Wiley & Sons: NY.
123. Otto, S., R.L. Furlan, and J.K. Sanders, *Recent developments in dynamic combinatorial chemistry.* Curr Opin Chem Biol, 2002. **6**(3): p. 321-7.
124. Huc, I. and R. Nguyen, *Dynamic combinatorial chemistry.* Comb Chem High Throughput Screen, 2001. **4**(1): p. 53-74.
125. Huc, I. and J.M. Lehn, *Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly.* Proc Natl Acad Sci U S A, 1997. **94**(6): p. 2106-10.
126. Bunyapaiboonsri, T., et al., *Dynamic deconvolution of a pre-equilibrated dynamic combinatorial library of acetylcholinesterase inhibitors.* Chembiochem, 2001. **2**(6): p. 438-44.
127. Cousins, G.R.L.P., S. -A.; Sanders, J. K. M., *Dynamic combinatorial libraries of pseudo-peptide hydrazone macrocycles.* Chemical Communications, 1999(16): p. 1575-1576.
128. Bunyapaiboonsri, T., et al., *Generation of bis-cationic heterocyclic inhibitors of Bacillus subtilis HPr kinase/phosphatase from a ditopic dynamic combinatorial library.* J Med Chem, 2003. **46**(26): p. 5803-11.
129. Hochgurtel, M., et al., *Target-induced formation of neuraminidase inhibitors from in vitro virtual combinatorial libraries.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3382-7.
130. Valade, A., D. Urban, and J.M. Beau, *Two galactosyltransferases' selection of different binders from the same uridine-based dynamic combinatorial library.* J Comb Chem, 2007. **9**(1): p. 1-4.
131. Zameo, S., B. Vauzeilles, and J.M. Beau, *Dynamic combinatorial chemistry: lysozyme selects an aromatic motif that mimics a carbohydrate residue.* Angew Chem Int Ed Engl, 2005. **44**(6): p. 965-9.
132. Valade, A., D. Urban, and J.M. Beau, *Target-assisted selection of galactosyltransferase binders from dynamic combinatorial libraries. An unexpected solution with restricted amounts of the enzyme.* Chembiochem, 2006. **7**(7): p. 1023-7.
133. Hioki, H.S., W. C., *Chemical Evolution: A Model System That Selects and Amplifies a Receptor for the Tripeptide (D)Pro(L)Val(D)Val.* Journal of Organic Chemistry, 1998. **63**(4): p. 904-905.
134. Otto, S.F., R. L. E.; Sanders J. K. M., *Dynamic Combinatorial Libraries of Macrocyclic Disulfides in Water.* Journal of the American Chemical Society, 2000. **122**(48): p. 12063-12064.
135. Ramstrom, O. and J.M. Lehn, *In situ generation and screening of a dynamic combinatorial carbohydrate library against concanavalin A.* Chembiochem, 2000. **1**(1): p. 41-8.
136. Milanesi, L., et al., *Amplification of bifunctional ligands for calmodulin from a dynamic combinatorial library.* Chemistry, 2006. **12**(4): p. 1081-7.
137. Hoerter, J.M., et al., *Discovery and mechanistic study of Al(III)-catalyzed transamidation of tertiary amides.* J Am Chem Soc, 2008. **130**(2): p. 647-54.

138. Eldred, S.E., et al., *Catalytic transamidation under moderate conditions.* J Am Chem Soc, 2003. **125**(12): p. 3422-3.
139. Bell, C.M., et al., *Catalytic metathesis of simple secondary amides.* Angew Chem Int Ed Engl, 2007. **46**(5): p. 761-3.
140. Cheeseman, J.D., et al., *Amplification of screening sensitivity through selective destruction: theory and screening of a library of carbonic anhydrase inhibitors.* J Am Chem Soc, 2002. **124**(20): p. 5692-701.
141. Corbett, A.D., et al., *Pseudodynamic combinatorial libraries: a receptor-assisted approach for drug discovery.* Angew Chem Int Ed Engl, 2004. **43**(18): p. 2432-6.
142. Shi, B., et al., *Discovery of glutathione S-transferase inhibitors using dynamic combinatorial chemistry.* J Am Chem Soc, 2006. **128**(26): p. 8459-67.
143. Shi, B. and M.F. Greaney, *Reversible Michael addition of thiols as a new tool for dynamic combinatorial chemistry.* Chem Commun (Camb), 2005(7): p. 886-8.
144. Duret, L., E. Gasteiger, and G. Perriere, *LALNVIEW: a graphical viewer for pairwise sequence alignments.* Comput Appl Biosci, 1996. **12**(6): p. 507-10.
145. Vuilleumier, S., *Bacterial glutathione S-transferases: what are they good for?* J Bacteriol, 1997. **179**(5): p. 1431-41.
146. Allocati, N., et al., *Glutathione transferases in bacteria.* Febs J, 2009. **276**(1): p. 58-75.
147. Frova, C., *Glutathione transferases in the genomics era: new insights and perspectives.* Biomol Eng, 2006. **23**(4): p. 149-69.
148. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.
149. Kiefer, F., et al., *The SWISS-MODEL Repository and associated resources.* Nucleic Acids Res, 2009. **37**(Database issue): p. D387-92.
150. Xiao, G., et al., *First-sphere and second-sphere electrostatic effects in the active site of a class mu gluthathione transferase.* Biochemistry, 1996. **35**(15): p. 4753-65.
151. Krengel, U., et al., *Crystal structure of a murine alpha-class glutathione S-transferase involved in cellular defense against oxidative stress.* FEBS Lett, 1998. **422**(3): p. 285-90.
152. Rossjohn, J., et al., *Structures of thermolabile mutants of human glutathione transferase P1-1.* J Mol Biol, 2000. **302**(2): p. 295-302.
153. Notredame, C., D.G. Higgins, and J. Heringa, *T-Coffee: A novel method for fast and accurate multiple sequence alignment.* J Mol Biol, 2000. **302**(1): p. 205-17.
154. Chang, M., J.L. Bolton, and S.Y. Blond, *Expression and purification of hexahistidine-tagged human glutathione S-transferase P1-1 in Escherichia coli.* Protein Expr Purif, 1999. **17**(3): p. 443-8.
155. Singh, S.P., et al., *Membrane association of glutathione S-transferase mGSTA4-4, an enzyme that metabolizes lipid peroxidation products.* J Biol Chem, 2002. **277**(6): p. 4232-9.

156. Yassin, Z., et al., *Role of mutation Y6F on the binding properties of Schistosoma japonicum glutathione S-transferase.* Int J Biol Macromol, 2003. **32**(3-5): p. 67-75.
157. Alin, P., U.H. Danielson, and B. Mannervik, *4-Hydroxyalk-2-enals are substrates for glutathione transferase.* FEBS Lett, 1985. **179**(2): p. 267-70.
158. Johansson, A.S., et al., *Structure-activity relationships and thermal stability of human glutathione transferase P1-1 governed by the H-site residue 105.* J Mol Biol, 1998. **278**(3): p. 687-98.
159. Biosciences, A. *GST Gene Fusion System Handbook*. 15/05/2010 [cited; Available from: http://www.gelifesciences.com/aptrix/upp01077.nsf/Content/Products?OpenDocument&moduleid=7336.
160. Nanduri, B. and P. Zimniak, *Role of active-site residues 107 and 108 of glutathione S-transferase mGSTA4-4 in determining the catalytic properties of the enzyme for 4-hydroxynonenal.* Arch Biochem Biophys, 1999. **362**(1): p. 167-74.
161. Hearne, J.L. and R.F. Colman, *Catalytically active monomer of class mu glutathione transferase from rat.* Biochemistry, 2006. **45**(19): p. 5974-84.
162. Zimniak, P., et al., *Naturally occurring human glutathione S-transferase GSTP1-1 isoforms with isoleucine and valine in position 104 differ in enzymic properties.* Eur J Biochem, 1994. **224**(3): p. 893-9.
163. Thompson, L.C., et al., *Double mutation at the subunit interface of glutathione transferase rGSTM1-1 results in a stable, folded monomer.* Biochemistry, 2006. **45**(7): p. 2267-73.
164. Johnson, W.W., et al., *Tyrosine 115 participates both in chemical and physical steps of the catalytic mechanism of a glutathione S-transferase.* J Biol Chem, 1993. **268**(16): p. 11508-11.
165. Parraga, A., et al., *The three-dimensional structure of a class-Pi glutathione S-transferase complexed with glutathione: the active-site hydration provides insights into the reaction mechanism.* Biochem J, 1998. **333 (Pt 3)**: p. 811-6.
166. Andujar-Sanchez, M., et al., *Thermodynamics of glutathione binding to the tyrosine 7 to phenylalanine mutant of glutathione S-transferase from Schistosoma japonicum.* Int J Biol Macromol, 2003. **32**(3-5): p. 77-82.
167. Bhat, V.T., et al., *Nucleophilic catalysis of acylhydrazone equilibration for protein-directed dynamic covalent chemistry.* Nat Chem. **2**(6): p. 490-7.
168. Copeland, R.A., *Evaluation of enzyme inhibitors in drug discovery*, ed. Wiley-interscience. 2005, Hoboken: John Wiley & Sons.
169. Chern, M.K., et al., *Tyr115, gln165 and trp209 contribute to the 1, 2-epoxy-3-(p-nitrophenoxy)propane-conjugating activity of glutathione S-transferase cGSTM1-1.* J Mol Biol, 2000. **300**(5): p. 1257-69.

Chapter 2: Probing the active site of a bacterial glutathione transferase using dynamic combinatorial chemistry

2.1. Introduction

2.1.1. Bacterial glutathione transferases

Bacterial GSTs have been studied since GST activity was reported in *Escherichia coli* in 1981 [1]. Since then, four different bacterial GST classes have been identified: Beta, Chi, Theta and Zeta. Most of them show poor catalytic activity and low specificity towards GST substrates such as CDNB. It has been proposed that these GSTs were not primarily involved in detoxification but were rather involved in primary metabolism. Their function may have changed during the course of evolution [1, 2, 3].

Bacterial theta class GSTs share high sequence similarity with eukaryotic theta class GSTs and the same properties. They are represented by dichloromethane (DCM) dehydrogenases which catalyse the glutathione dependant conversion of DCM to formaldehyde and allow facultative methylotrophic bacteria to use DCM as a carbon source. A tetrachlorohydroquinone (TCHQ) dehalogenase has been reported as belonging to the zeta class, based on amino-acid sequence alignments. This enzyme is involved in the degradation of pentachlorophenol and has isomerase activity like the eukaryotic zeta class GSTs. Beta-class GSTs have been isolated from several bacteria and are unusual in that they have two catalytic cysteine residues in their G-site. They are able to bind GSH-affinity matrices and they can conjugate GSH to CDNB. In contrast a new GST class called 'Chi' was isolated recently from cyanobacteria and they lack the two cysteine residues in their active site [1, 2, 3, 4, 5, 6].

2.1.2. The beta-class glutathione transferase BphK

The bacterial GST BphK (Figure 2.1.) was expressed from a gene found in the genome of the species *Burkholderia xenovorans* LB400. This species is an environmental strain that was first isolated from a landfill site in the USA known to

contain a large concentration of polychlorinated biphenyls (PCBs) [7, 8]. It has a remarkable ability to detoxify PCBs, via the detoxification pathway known as the Bph pathway, which it then proceeds to use as a carbon source. The Bph pathway consists of a number of enzymes that catabolise PCBs. *B. xenovorans*' ability to detoxify PCBs has been of great importance to the research community as a potential candidate for microbial degradation of pollutants [9].

The Bph pathway is coded by the Bph operon in the genome of *B. xenovorans*. This operon consists of a cluster of genes that code for enzymes that perform the various stages of the biphenyl catabolism eventually leading to the production of acetyl-CoA. When the Bph operon was characterised in detail the unexpected presence of a gene (*bphK*) coding for the GST enzyme mentioned above was observed. Although its function was unclear it was found not to be essential for growth of the bacteria on biphenyls [10].

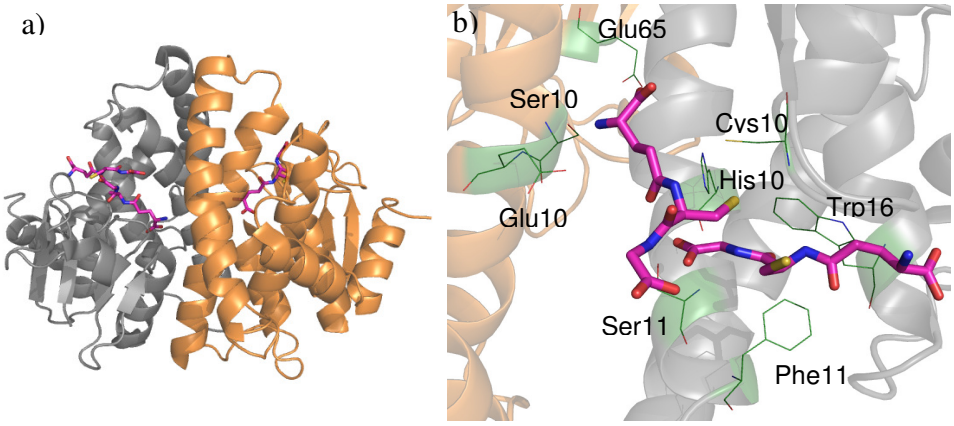


Figure 2.1. Crystal structure of BphK from *B. xenovorans* (PDB 2GDR): a) homodimer (one subunit in grey, the other in orange) in complex with three molecules of GSH (shown in pink); b) active site residues (in green) interacting with the two molecules of GSH [157].

Subsequent studies have proposed that this GST may be responsible for the dechlorination of the chlorinated 2-hydroxy-6-oxo-6-phenyl-2,4-dienoate (HOPDA) metabolites of the PCB catabolic pathway (Figure 2.2.) [11, 12, 13]. In the proposed two-step mechanism for BphK's function, two molecules of GSH are required. In the first half of the reaction, 3Cl-HOPDA is dechlorinated. In the second half, the

catalytically active cysteine (Cys10) and the molecule of GSH used in the first step are regenerated.

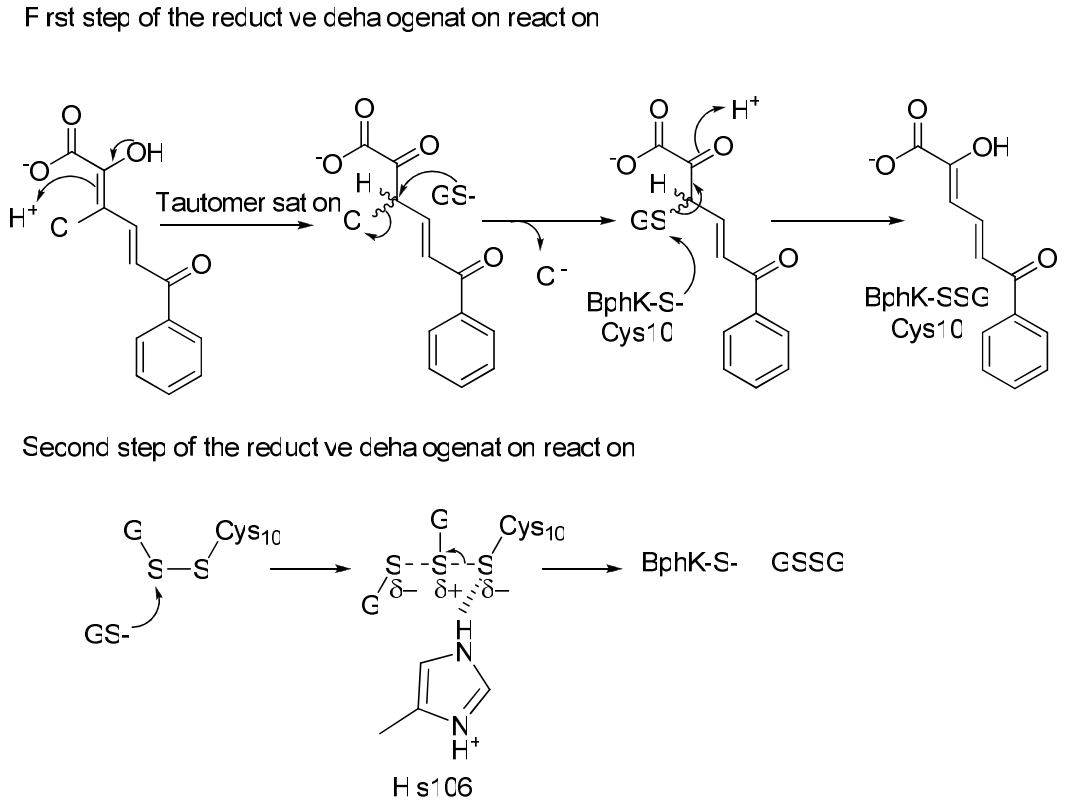


Figure 2.2. Proposed mechanism for the dehalogenation of 3-Cl HOPDA by BphK [15].

Subsequent searches on genomes of other bacteria resulted in the discovery of GST proteins with high sequence identity to that of BphK. Many bacteria of the genus *Burkholderia*, *Pseudomonas* and *Sphingomonas* have been shown to have proteins sharing sequence identity to the BphK GST [15]. *B. cenocepacia* belongs to the *Burkholderia cepacia* complex (Bcc) which is used to describe some nine species of the genus *Burkholderia*. Members of the Bcc are known to exist in the environment and are also pathogens of both humans and plants. They are most notorious as pathogens in sufferers of cystic fibrosis (CF) where they are opportunistic colonisers of the patient's lungs [14].

The J2315 strain of the ET-12 lineage of *Burkholderia cenocepacia* was isolated from the lungs of CF patients. It is a known member of the Bcc and has spread among patients in Edinburgh, Scotland and Toronto, Canada during an epidemic outbreak in the 1990's. As well as being able to survive in the CF lung *B.cenocepacia* inhabits the rhizosphere and has an impressive ability to detoxify hostile environments [14]. Studies have shown that the J2315 strain of *B. cenocepacia*, is capable of surviving murine macrophages, where a mechanism is in place that delays the maturation of phagosomes [15]. This ability to escape host cell's fundamental immune responses along with J2315's known occurrence in CF patients lungs shows conclusively that is has a very impressive detoxification system.

2.1.3. Aims and research strategy

BphK isolated from *B. xenovorans* is one of the most extensively investigated bacterial glutathione transferase for bioremediation since it is capable of degrading a broad spectrum of polychlorinated biphenyls (PCBs). It has an interesting mechanism of reaction which differs from the other GSTs by using two molecules of GSH instead of one.

A bacterial strain from the same genus as *B. xenovorans*, *B. cenocepacia* J2315 can infect CF patients and is resistant to antibiotics. As a detoxification and PCB degrading enzyme, BphK from J2315 may be partly responsible for the bacteria's resistance to a highly hostile environment. Therefore, the aims of this project were to:

- search the genome of *B. cenocepacia* J2315 for a gene similar to the BphK gene found in *B. xenovrans*;
- isolate the gene from J2315;
- express the new GST protein;
- characterise the activity of the BphK_J2315 protein with a CDNB assay;
- probe the enzyme active site with a DCC approach.

2.2. Results and Discussion

2.2.1. BphK_J2315 isolation and purification

The BphK protein sequence from *B. xenovorans* strain LB400 was used to search the genome of *B. cenocepacia* for a similar translated sequences and a match of 45% identity was observed with the translated gene BCAM0431. Importantly, the active site residues were conserved and the C-terminal motif Ser/Thr-Xaa-Xaa-Asp common to all GSTs from PCB degrading operons was present (Figure 2.3.). Site-directed mutagenesis of this motif in LB400 BphK affected GST activity towards the substrates CDNB and 4-chlorobenzoate (4-CBA) [16].

```
               10        20        30        40        50        60
BphK   MKLYYSPGACSLSPHIALREAGLNFELVQVDLASKKTASGQDYLEINPAGYVPCLQLDDG
       :::::.::.:: .  :.:::: .. :.:.::  .. . .:.:: ..:  :::: :.::::
BphK_J MKLYHAPGSCSQAICIVLREADIDAEIVKVDARKHVVDGGRDYYDVNELGYVPLLELDDG
               10        20        30        40        50        60

               70        80        90       100       110       120
BphK   RTLTEGPAIVQYVADQVPGKQLAPANGSFERYHLQQWLNFISSELHKSFSPLFNPASSDE
         : :::.:.::.:::.:   :::: :.. ::.:.::::...:.::.: ::.  ... .
BphK_J TLLREGPVIAQYLADQLPEAALAPAYGTLARYRLMEWLNFLGTEIHKGFIPLLYAVQAGK
               70        80        90       100       110       120

              130       140       150       160               170
BphK   WKNAVRQSLNTRLGQVARQLEHAPYLLGDQLSVADIYLFVVLGW-------SAY-VNIDL
       . . :::.:. :.. . :::.   .: :: ..::: ::::. ::       :.: ..:::
BphK_J YVEPVRQKLDGRFAWIDRQLDGRTFLTGDTFTVADAYLFALTGWGKADWMRSVYNADIDL
              130       140       150       160       170       180

            180       190       200
BphK   SPWPSLQAFQGRVGGREAVQSALRAEGLIK
       :    :.:.  ::  : ::::.: :..:..
BphK_J SRHAHLRAWYERVRERPAVQAVLAADNLLR
              190       200       210
```

Figure 2.3. Alignment of the protein sequences of BphK from *B. xenovorans* (BphK) with the translation of the gene BCAM0431 from *B. cenocepacia* J2315 (BphK_J). Alignment was made with LALIGN and gave 45.2% of identity between the two sequences. The conserved catalytic residues are shown in red; the conserved C-terminal motif is highlighted in yellow.

The gene was isolated from the *B. cenocepacia* J2315 genome, sequenced, and cloned into the expression plasmid pET22b as described in section 5.2.4.9.5. The protein of 23769 Da was named BphK_J2315 and expressed in *E. coli* BL21 (DE3) at 37 ˚C. After 3 hours of induction with IPTG, a clear expression of the protein could be observed by SDS-PAGE analysis (Figure 2.4.).

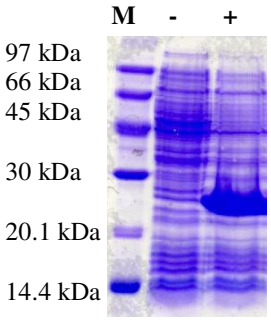


Figure 2.4. SDS-PAGE showing expression of BphK_J2315 from *E. coli* before (-) and after (+) induction with IPTG. From comparison with the marker (M) bands it is possible to see a large increase in the concentration of proteins of approximately 24 kDa.

Purification of BphK_J2315 was performed by affinity chromatography, using a 20 ml GSTPrep FF 16/10 column (Figure 2.5.), followed by size exclusion chromatography using a 360 ml Superdex S200 column (Figure 2.6.). Specific activities were measured at each purification step, using a CDNB assay.

Purification and characterization of BphK_J2315

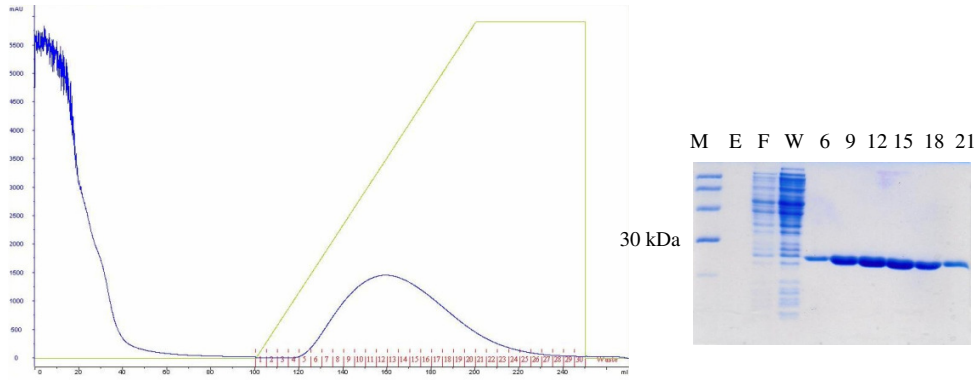


Figure 2.5. Elution profile of BphK_J2315 from the GSTPrep column showing the A_{280}. The collected fractions were analysed by SDS-PAGE (M: low molecular weight marker; E: cell free extract; F: flow through; W: wash; collected fractions are represented by their number).

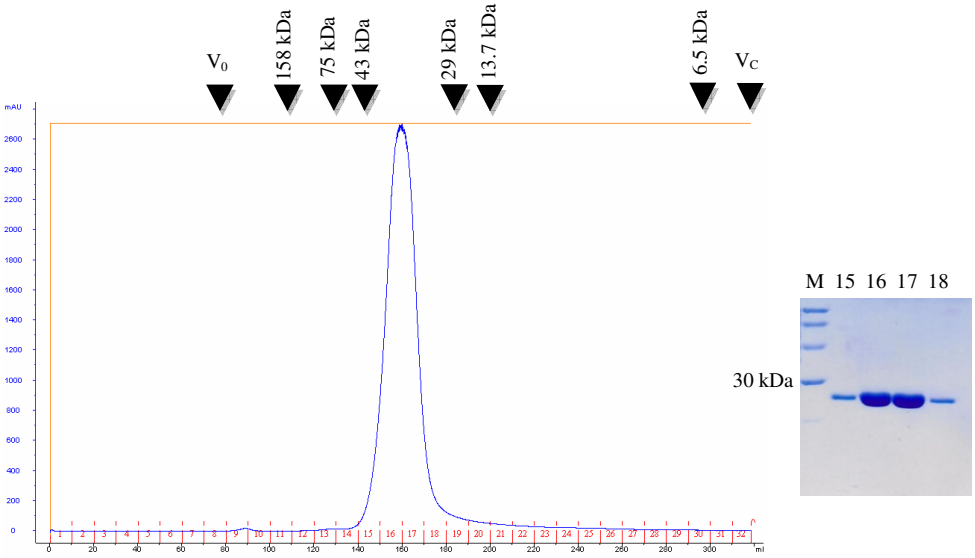


Figure 2.6. Elution profile of BphK_J2315 from the Sephacryl S-200 column showing the A_{280}. Calibration markers are specified at the top of the S-200 UV trace, V_0 is a void volume, V_C is the column volume. Collected fractions were analysed by SDS-PAGE (M: low molecular weight marker).

Table 2.1. Purification of BphK_J2315

	Volume (ml)	Total protein (mg)	[Protein] (mg/ml)	Total activity (µmol/min/ml)	Specific activity (µmol/min/mg)	Yield (%)
Cell free extract	50	336	6.7	16.8 ± 1.8	2.5 ± 0.3	100
S-200 eluate	40	104	2.6	11.4 ± 1.2	4.4 ± 0.5	68

A yield of ~ 100 mg of recombinant protein was obtained from 3 liters of *E. coli* BL21(DE3) culture. The molecular mass of native BphK_J2315 was estimated to be 40 kDa by calibrated gel filtration chromatography, which is close to the dimeric molecular mass of ~ 47.5 kDa. Murphy and co-workers purified BphK from *Burkholderia xenovorans* strain LB400 with a glutathione affinity resin and determined a mass of approximately 33.9 kDa for the native enzyme. Crystallographic analyses revealed that the enzyme is dimeric [11]. BphK_J2315 was purified to its complete homogeneity with a total recovery of 68 % of the total activity using the CDNB assay. The specific activity of the purified enzyme towards the reporter substrate CDNB, 4.4 μmol/min/mg, was 1.8-fold higher than that of the raw extract. A similar specific activity of 7.4 μmol/min/mg was found for BphK from *Burkholderia xenovorans* strain LB400 [13].

In order to estimate the purity of the sample eluted from the gel filtration column and to determine the mass of the purified GST, MALDI-TOF MS analysis was carried on after dialysis against a 20 mM Tris/HCl buffer. Two major fused peaks of 23,761.68 and 24,076.09 m/z were observed (Figure 2.7.). The first peak corresponds to the BphK_J2315 monomer which has a theoretical mass of 23,769 Da. The second peak has an additional 314 Da which is relatively close to the mass of glutathione (307 Da). Since BphK is known to have a catalytically active cysteine in its active site which binds GSH covalently, we can assume that this peak corresponds to BphK_J2315 bound to one molecule of glutathione. This hypothesis is verified by the appearance of a third peak at 48,145.13 m/z, matching with the molecular weight of the dimeric form of BphK_J2315 (47,538 Da) plus two molecules of GSH (614 Da).

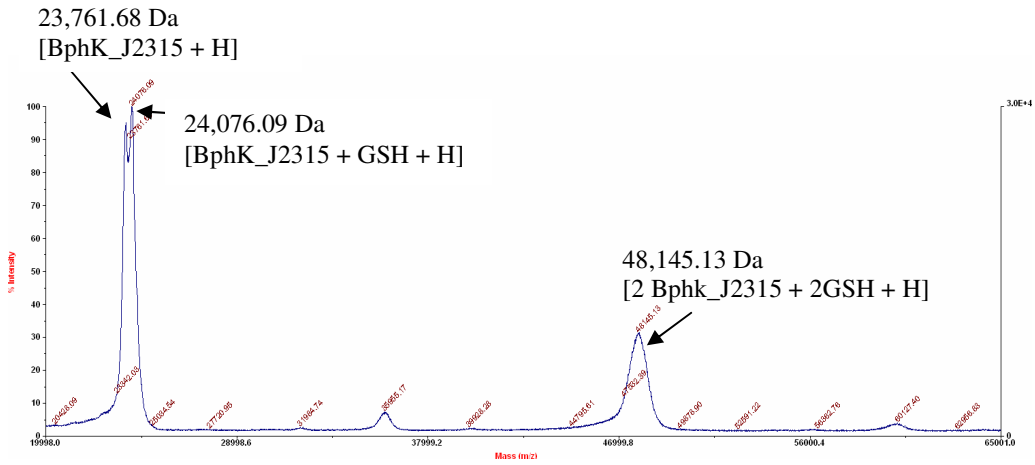


Figure 2.7. MALDI-TOF mass spectrum of BphK_J2315 from 20,000 to 65,000 m/z.

Steady-state kinetics with variable concentrations of substrates (GSH: 1.0 µM to 2.0 mM; CDNB: 0.5 µM to 1.0 mM) were performed. Data were plotted using the Michaelis-Menten and Lineweaver-Burk equations, and kinetic parameters such as k_{cat} and K_M were determined for each substrate. k_{cat} was calculated using the monomeric molecular weight. The same results were obtained with the two models; Michaelis-Menten plots are shown in Figure 2.8. and kinetic data are summarised in Table 2.1.

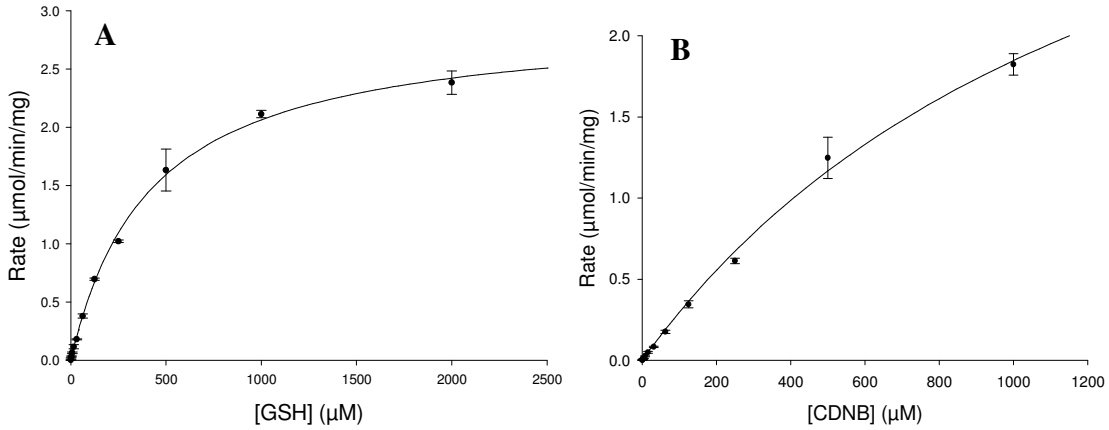


Figure 2.8. Michaelis-Menten representation of BphK_J2315 activity towards the substrates GSH (**A**) and CDNB (**B**). GSH and CDNB concentrations were varied from 1 µM to 2 mM and 0.5 µM to 1 mM, respectively.

Table 2.1. Kinetic parameters of BphK_J2315

	K_M (mM)	V_{max} (µmol/min/mg)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}.s^{-1}$)
GSH	0.42 ± 0.03	2.92 ± 0.31	1.15 ± 0.12	2.74
CDNB	1.40 ± 0.23	4.43 ± 0.49	1.75 ± 0.19	1.25

K_M and k_{cat} for BphK from *B. xenovorans* has not been reported in the literature. However, similar data were obtained for another beta class GST from *Proteus mirabilis*, GST B1-1, which has K_M values of 0.69 and 0.73 mM and k_{cat} of 0.97 and 1.16 towards GSH and CDNB respectively [17].

Since BphK_J2315 was able to conjugate GSH to CDNB, we decided to probe its active site using the CDNB-based DCL described earlier in Chapter 1.

2.2.2. Exploration of BphK_J2315's active site using DCC

Venughopal Bhat interfaced the glutathione-conjugated 10-membered library presented in Figure 2.9. with the enzyme for 16h and measured the amplifications (Figure 2.10.).

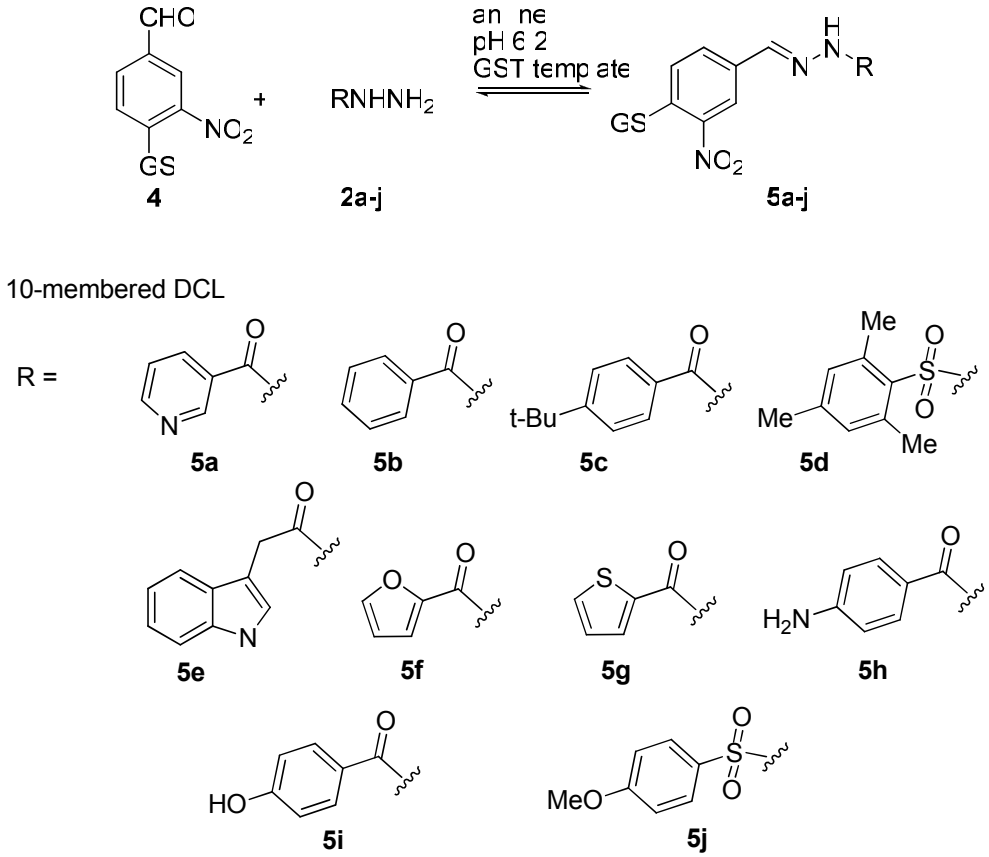


Figure 2.9. Aniline-catalysed acyl hydrazone formation. Conditions: Aldehyde (5 μM), hydrazides (20 μM) in NH_4OAc Buffer (50 mM, pH 6.2) containing 15% DMSO. Library A is run in the presence of aniline (20mM).

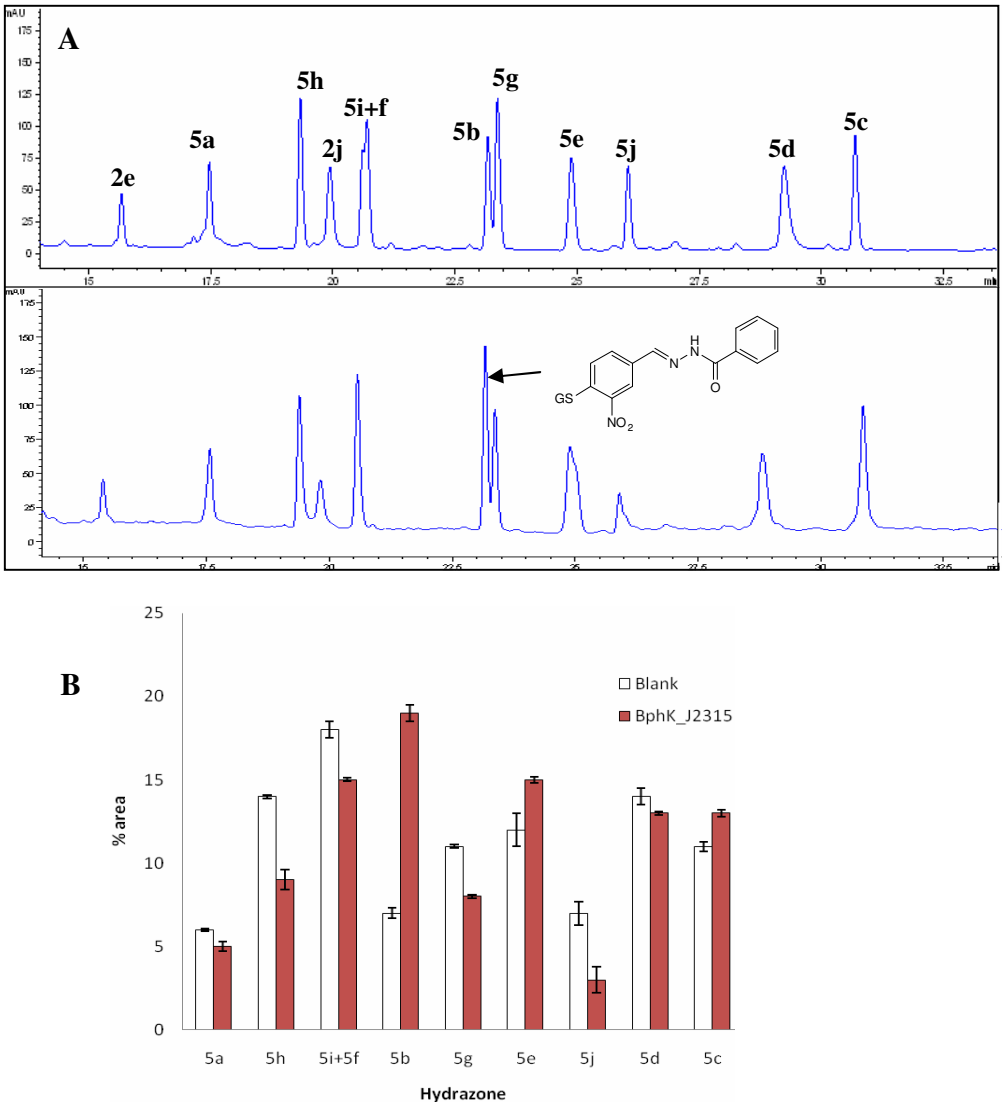


Figure 2.10. HPLC analyses of BphK_J2315-templated DCL of GSH conjugates. **A**, DCL acyl hydrazone composition in the absence of target (blank) and in the presence of BphK_J2315. DCL conditions: GST (1 equiv.), aldehyde (5 μM), hydrazides (20 μM) and aniline (10 mM) in NH_4OAc buffer (50 mM, pH 6.2) containing 15% DMSO for 16 h. **B**, Changes in DCL component concentration for blank and BphK_J2315 DCLs.

The ten peaks corresponding to the 10 hydrazones were identified in both the blank and the GST-templated DCLs. Three compounds **5b**, **5e** and **5c** were found to be amplified. The strongest increase in peak area was observed for hydrazone **5b**

which was amplified to almost 200% as compared to **5e** (~20%) and **5c** (~15%). The seven remaining members decreased in proportion. The three selected ligands are all hydrophobic. The hydrazone **5b** shares some structural features with the well-known substrate of BphK, HOPDA (Figure 2.11.). In particular, both molecules are planar and have a benzene ring at one extremity. Murphy and co-workers have crystallised BphK from LB400 in complex with glutathione and HOPDA [11]. These structures revealed the presence of a solvent-exposed H-site, where the benzene ring of HOPDA makes hydrophobic contacts with Trp164, Tyr167 and Phe113. These three residues are conserved in BphK_J2315. Therefore, one can hypothesise that the benzene ring of **5b** is accommodated in the corresponding hydrophobic pocket of BphK_J2315. Similarly, **5e** and **5c**, which also have hydrophobic aromatic rings, may be able to bind to the same region with a lower affinity.

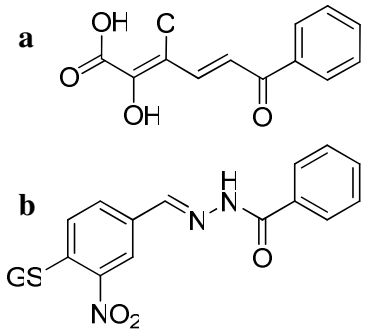


Figure 2.11. Structure of HOPDA (**a**) and **5b** (**b**).

2.3. Conclusions

A BphK enzyme was isolated from the *B. cenocepacia* genome, expressed in *E. coli* and purified using glutathione-affinity chromatography. Kinetic parameters were determined and were comparable to the data found for some other bacterial beta class GSTs. They also showed that BphK_J2315 is able to conjugate CDNB to glutathione. The enzyme successfully amplified three hydrazones from the glutathione-conjugated DCL. The highest amplified member was found to share some structural features with a known substrate of BphK and could possibly bind to a hydrophobic pocket located in the active site of the enzyme. Although these results strengthen the credibility of the use of DCC as a tool in drug discovery, structural analyses of the complex BphK:**5b/5e/5c** as well as inhibition studies remain to be done to provide insight into the protein/ligand interactions involved and a rationale for the observed selectivity.

2.4. Chapter 2 references

1. Shishido, T., *Glutathione S-Transferase from Escherichia coli.* Agriculture Biological Chemistry, 1981. **45**(12): p. 2951-2953.
2. Vuilleumier, S., *Bacterial glutathione S-transferases: what are they good for?* J Bacteriol, 1997. **179**(5): p. 1431-41.
3. Vuilleumier, S. and M. Pagni, *The elusive roles of bacterial glutathione S-transferases: new lessons from genomes.* Appl Microbiol Biotechnol, 2002. **58**(2): p. 138-46.
4. Allocati, N., et al., *Glutathione transferases in bacteria.* FEBS J, 2009. **276**(1): p. 58-75.
5. Sheehan, D., et al., *Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily.* Biochem J, 2001. **360**(Pt 1): p. 1-16.
6. Wiktelius, E. and G. Stenberg, *Novel class of glutathione transferases from cyanobacteria exhibit high catalytic activities towards naturally occurring isothiocyanates.* Biochem J, 2007. **406**(1): p. 115-23.
7. Goris, J., et al., *Classification of the biphenyl- and polychlorinated biphenyl-degrading strain LB400T and relatives as Burkholderia xenovorans sp. nov.* Int J Syst Evol Microbiol, 2004. **54**(Pt 5): p. 1677-81.
8. Hofer, B., S. Backhaus, and K.N. Timmis, *The biphenyl/polychlorinated biphenyl-degradation locus (bph) of Pseudomonas sp. LB400 encodes four additional metabolic enzymes.* Gene, 1994. **144**(1): p. 9-16.
9. Pieper, D.H. and M. Seeger, *Bacterial metabolism of polychlorinated biphenyls.* J Mol Microbiol Biotechnol, 2008. **15**(2-3): p. 121-38.
10. Bartels, F., et al., *Occurrence and expression of glutathione-S-transferase-encoding bphK genes in Burkholderia sp. strain LB400 and other biphenyl-utilizing bacteria.* Microbiology, 1999. **145 (Pt 10)**: p. 2821-34.
11. Tocheva, E.I., et al., *Structures of ternary complexes of BphK, a bacterial glutathione S-transferase that reductively dechlorinates polychlorinated biphenyl metabolites.* J Biol Chem, 2006. **281**(41): p. 30933-40.
12. Gilmartin, N., et al., *BphK shows dechlorination activity against 4-chlorobenzoate, an end product of bph-promoted degradation of PCBs.* FEMS Microbiol Lett, 2003. **222**(2): p. 251-5.
13. Fortin, P.D., et al., *A glutathione S-transferase catalyzes the dehalogenation of inhibitory metabolites of polychlorinated biphenyls.* J Bacteriol, 2006. **188**(12): p. 4424-30.
14. Mahenthiralingam, E., T.A. Urban, and J.B. Goldberg, *The multifarious, multireplicon Burkholderia cepacia complex.* Nat Rev Microbiol, 2005. **3**(2): p. 144-56.
15. Lamothe, J., et al., *Intracellular survival of Burkholderia cenocepacia in macrophages is associated with a delay in the maturation of bacteria-containing vacuoles.* Cell Microbiol, 2007. **9**(1): p. 40-53.

16. Gilmartin, N., D. Ryan, and D.N. Dowling, *Analysis of the C-terminal domain of Burkholderia sp. strain LB400 BphK reveals a conserved motif that affects catalytic activity.* FEMS Microbiol Lett, 2005. **249**(1): p. 23-30.
17. Perito, B., et al., *Molecular cloning and overexpression of a glutathione transferase gene from Proteus mirabilis.* Biochem J, 1996. **318 (Pt 1)**: p. 157-62.

Chapter 3: Preliminary work on the exploration of a glutathione transferase-homolog using dynamic combinatorial chemistry

3.1. Introduction

3.1.1. Eukaryotic translation elongation factor 1Bγ(eEF1Bγ) from *Saccharomyces cerevisiae*

Protein biosynthesis is divided into three phases: initiation, elongation and termination. During initiation, a Met-tRNA$_i^{Met}$ is based paired with a starting AUG codon of an mRNA within the ribosome. In the elongation phase, one amino-acid is added at a time to a growing polypeptide; the ribosome selects the aminoacylated-tRNA (aa-tRNA) according to the sequence of codons in the mRNA. This process is assisted by several protein elongation factors. In eukaryotes, aa-tRNAs are brought to the ribosome by eukaryotic elongation factor 1A (eEF1A), in the form of a ternary complex aa-tRNA-eEF1A-GTP. Upon cognate recognition, GTP is hydrolysed to GDP and eEF1A leaves the ribosome. Finally, nucleotide exchange of eEF1A is catalysed by the eukaryotic elongation factor 1B (eEF1B), so that inactive eEF1A-GDP is converted into active eEF1A-GTP which can then bring another aa-tRNA to the ribosome [1].

In *Saccharomyces cerevisiae*, eEF1B contains three subunits, α, β and γ. The last two interact with each other through their N-termini [2]. α and β are the catalytic subunits in the guanine exchange process [3]. eEF1Bγ has been shown to bind tubulin [4], the membrane of the endoplasmic reticulum [5] and mRNA [6] and is therefore thought to play a role in directing the protein synthetic apparatus towards cytoplasmic structures in the cell. Two eEF1Bγ isoforms, Tef3p and Tef4p, have been isolated from *Saccharomyces cerevisiae* [7]. Although its function is still unclear, Tef3p has been identified as a calcium-dependant membrane binding protein and as a gene dosage suppressor of a 40S ribosomal subunit assembly mutant [8]. It is also suspected to play a role in the oxidative stress pathway [9] and in the regulation of expression of methionine sulfoxide reductase [10].

Valencia and co-workers have shown by computational methods that the N-terminal domain of eEF1Bγ has sequence motifs characteristic to the theta class GSTs [11]. GST activity has been detected in rice eEF1Bγ [12] and a related trypanothione transferase activity has been ascribed to eEF1Bγ from the protozoa *Leishmania major* [13]. Yeast eEF1Bγ consists of two structural domains connected by a flexible peptide linker [2, 3]. The C-terminal domain 2 is monomeric and does not interact with eEF1Bα. The structure of the N-terminus of Tef3p from *S. cerevisiae* has been solved by Jeppesen *et al.* [3] and revealed a clear structural conservation to GST (Figure 3.1.). Since the full-length Tef3p tends to aggregate at physiological conditions, they constructed two deletion mutants, based on trypsin digestion and sequence alignment with the human theta class GST T2-2 sequence. The two resulting recombinant fragments, domain 1 (1-219 residues) and domain 1' (1-242 residues) respectively, had both sequence homology with GST and were highly soluble. Interestingly, domain 1 was found to be monomeric, whereas domain 1' was purified as a dimer, suggesting that residues 220-242 were required for dimerisation.

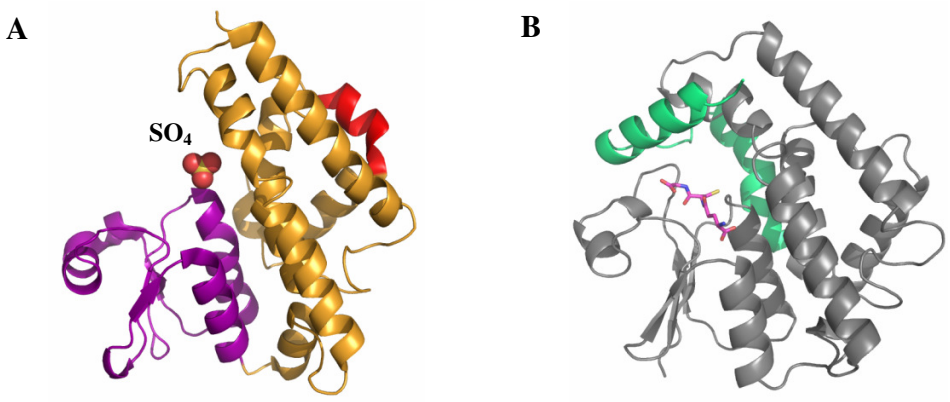


Figure 3.1. A, the eEF1Bγ domain 1 monomer viewed from the dimer interface side. The N-terminal subdomain 1_N containing the sulphate ion is shown in purple, the C-terminal subdomain 1_C in yellow, the extra-helix between α-6 and α-7 is shown in red. B, the human GST T2-2 monomer with GSH shown in pink bound in the active site [60]. Helices α-8 and α-9, shown in green, encapsulate the active site.

Only the crystal structure of domain 1 could be solved. Similar to other GST structures, domain 1 consists of 2 sub-domains connected by a short peptide linker with sub-domain 1_N adopting a topology similar to a thioredoxin fold and sub-domain 1_C containing a hydrophobic core created by interactions between the aromatic rings of Trp99, Phe143, Phe169, Phe173, Trp181, Trp191, and Phe192. Compared to the human GST T2-2 structure, there is an additional helix between α-6 and α-7 similar to that found in the *Arabidopsis thaliana* GST theta class structure. Domain 1 crystallised as a dimer, showing mainly hydrophobic and polar interactions between the two monomers, including a hydrophobic 'lock and key' motif described for mammalian GST structures. In the putative active site, several conserved residues involved in GSH binding were identified and a potential catalytic Tyr7, although orientated away from the active site, was pointed out. Despite these observations, domain 1 did not bind to GSH affinity resin nor did it show any activity towards CDNB *in vitro* [3].

It has been suggested that domain 1' of eEF1Bγ mediates the dimerisation of the eEF1 complex. Based on high sequence identity after alignment with other eEF1Bγ isoforms, an interaction site with eEF1Bα has been proposed to be located close to the putative GST active site. A communication may then occur between the exchange activity and the putative GST activity of eEF1Bγ. Moreover, it has been suggested that eEF1Bβ could interact with the loop located between α-2 and β-2, which would alternatively become part of the GST active site, giving it more flexibility and allowing substrate binding and catalysis. Although Tef3p did not show any GST activity towards CDNB, catalytic activity with a more specific secondary substrate was not excluded.

3.1.2. Aims and research strategy

eEF1Bγ is a protein homologous to GSTs, likely active, for which a potential secondary substrate remains to be characterised. Since its putative active site shares common structural features with CDNB-conjugating GSTs, the DCL created by Venugopal Bhat, based on the CDNB scaffold, was used to probe the "H-site" binding pocket in the hope of identifying an amplified ligand. Moreover, finding a ligand for the longer domain 1' could help for a more qualitative crystallisation of the dimer.

3.2. Results and Discussion

3.2.1. Expression and purification

Genes coding for domain 1 and domain 1' were both cloned in pET11-d by Dr. Andersen and co-workers and both contained an N-terminal 6-histidine tag [3]. The resulting theoretical molecular weights of the tagged proteins were 24,876.65 Da for domain 1 and 28,042.26 Da for the longer domain 1'. Plasmids were named pTKB588 and pTKB611 respectively.

The two proteins were expressed at 37 °C in *E. coli* BL21 (DE) host cells, using the expression vectors pTKB588 and pTKB611. After induction with IPTG for 5 hours, an increase in expression of the recombinant proteins of the expected masses (~25 and 28 kDa) was observed by SDS-PAGE analysis (Figure 3.2.).

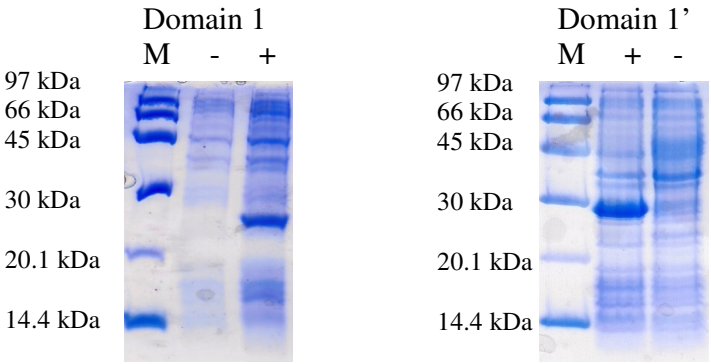


Figure 3.2. SDS-PAGE gels showing expression of proteins from *E. coli* before (-) and after (+) induction with IPTG. From comparison with the marker (M) bands it is possible to see increase in the concentration of proteins of approximately 25 and 28 kDa.

Domains 1 and 1' were purified using a nickel column, followed by gel filtration chromatography, a 120 ml Superdex S-75 column and a 360 ml Sephacryl S-200 column were used respectively. Cell free extracts, affinity columns flow-throughs (containing unbound proteins) and washes, and the collected fractions were used for SDS-PAGE analysis (Figures 3.3. and 3.4.).

Purification of domain 1

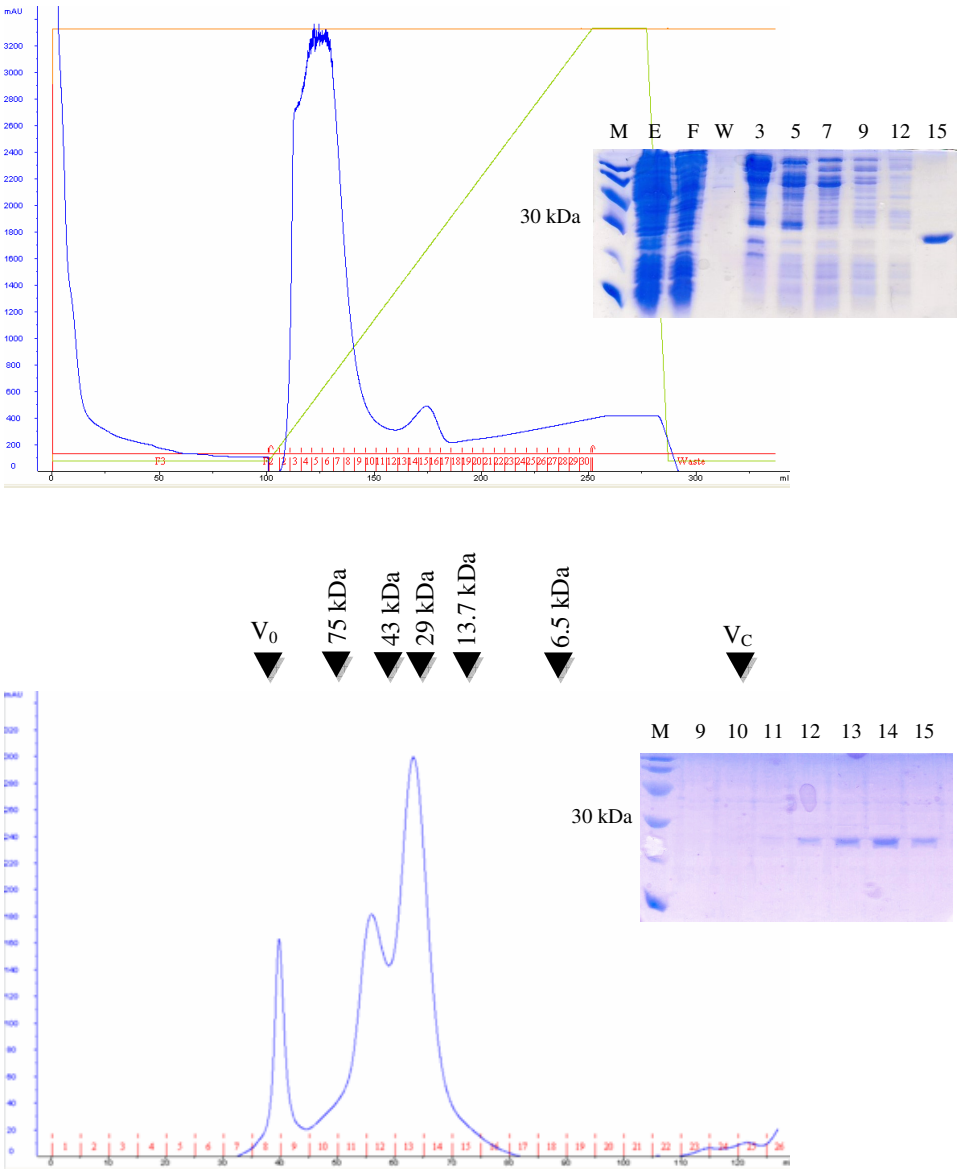


Figure 3.3. Elution profiles of domain 1 from the Ni column (A) and from the Superdex S-75 (B) showing the A_{280}. Calibration markers are specified at the top of the S-75 UV trace, V_0 is a void volume, V_C is the column volume. Collected fractions were analysed by SDS-PAGE (M: low molecular weight marker; E: cell free extract; F: flow through; W: wash; collected fractions are represented by their number).

Purification of domain 1'

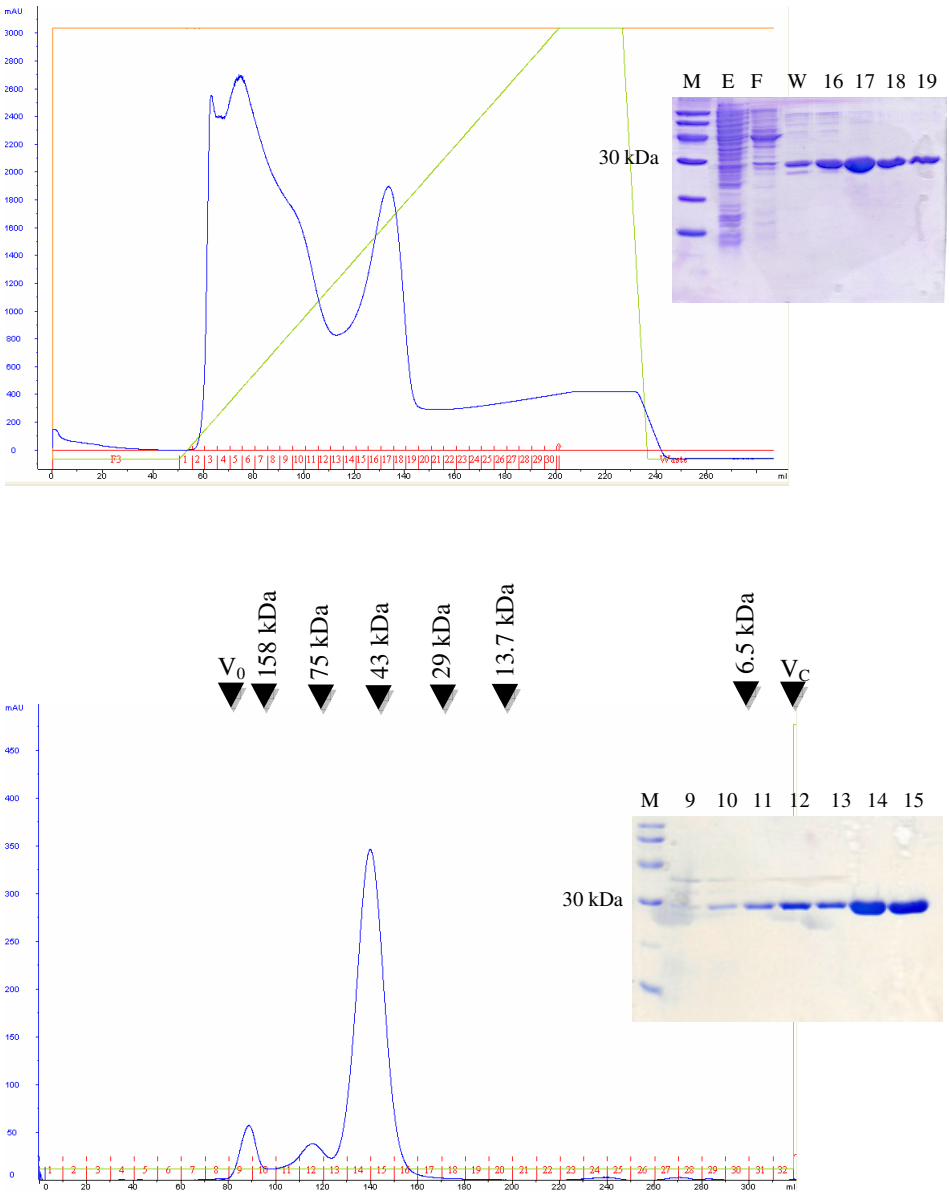


Figure 3.4. Elution profiles of domain 1' from the Ni column (A) and from the Sephacryl S-200 (B) showing the A_{280}. Calibration markers are specified at the top of the S-200 UV trace, V_0 is a void volume, V_C is the column volume. Collected fractions were analysed by SDS-PAGE.

UV-vis traces from the nickel columns showed large amounts of protein contaminants eluting from fraction 1 to fraction 12 in both purifications. Nickel is the most widely available metal ion for purifying histidine-tagged protein and generally provides good binding efficiency but also tends to bind non-specifically to endogenous proteins that contain histidine clusters. Jeppesen *et al*. used a cobalt resin to purify the same two proteins. Cobalt exhibits a more specific interaction with histidine-tags, resulting in a less non-specific interaction.

The molecular masses of the native proteins were estimated to be 25 kDa for domain 1 and 55 kDa for domain 1' by calibrated gel filtration chromatography, indicating that domain 1 is monomeric (theoretical mass of 24,876.65 Da) and domain 1' is dimeric (56,084 Da). A yield of ~ 40 µg was obtained for domain 1, which was not enough material for LC ESI-MS analysis and for protein directed DCC. Therefore, my work focused on domain 1' for which a yield of ~ 1 mg per litre of cell culture was obtained.

Eluted fractions from the Sephacryl S-200 containing domain 1' (F14 and 15) were pooled and dialysed against a 20 mM Tris/HCl buffer prior to analysis by MALDI-TOF mass spectrometry. The resulting spectrum is presented in figure 3.4. A fused peak showing two masses of 27,914.12 Da and 28,044.60 Da was observed. The first mass corresponded to the theoretical mass of domain 1' without one methionine plus one proton (27,912 Da), the second mass to domain 1' plus one proton (28,043 Da). These data suggest that a part of the proteins had their N-terminal methionine cleaved by *E. coli* during their synthesis. A third peak of 56,086.20 Da was relatively close to the dimeric theoretical molecular weight of domain 1' plus one proton (56,089 Da).

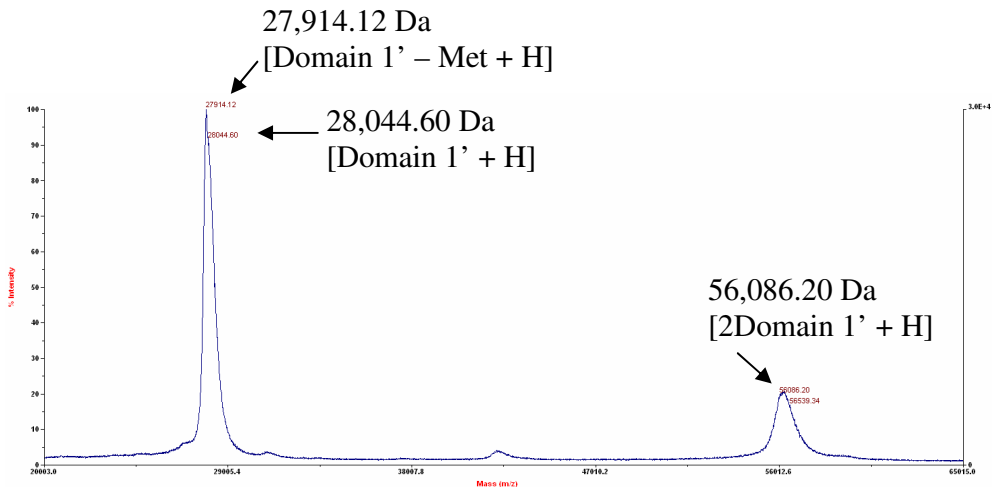


Figure 3.4. MALDI-TOF mass spectrum of domain 1' from 20,000 to 65,000 m/z.

3.2.2. Probing domain 1' using dynamic combinatorial chemistry

The glutathione-conjugated 10-membered library described in chapters 1 and 2 was used to probe domain 1'. Venugopal Bhat interfaced the DCL with the protein for 16h and compared the acyl hydrazone composition with that of the same DCL in the absence of the target protein, in order to identify any amplification (Figure 3.5.).

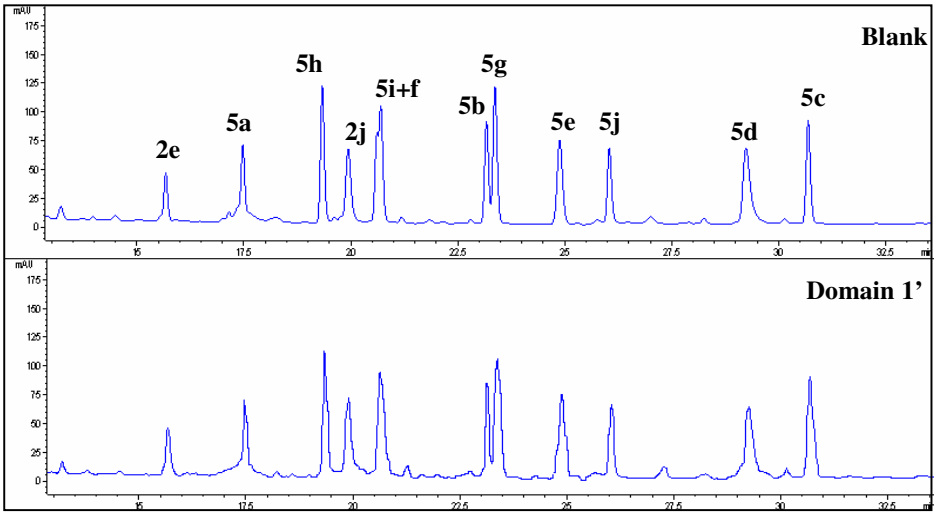


Figure 3.5. HPLC analyses of Domain 1'-templated DCL of GSH conjugates; DCL acyl hydrazone composition in the absence of target (blank) and in the presence of Domain 1'. DCL conditions: Domain 1' (1 equiv.), aldehyde (5 μM), hydrazides (20 μM) and aniline (10 mM) in NH_4OAc buffer (50 mM, pH 6.2) containing 15% DMSO for 16 h.

By comparing the two UV traces from the HPLC analyses, we observed that all the peaks remained unchanged in both chromatograms. Therefore, the presence of domain 1' did not promote any change in the library composition. Since CDNB is not a substrate for eEF1Bγ, it is possible that none of the members were able to bind the protein. Binding studies such as isothermal calorimetry will be necessary to confirm these results.

Since the putative GST active site of domain 1' resembles that of the human GST T2-2, I searched for the literature-known substrates for this enzyme in order to be able to design a more appropriate library for domain 1'. Interestingly, hGST T2-2 does not have any activity towards CDNB and does not bind to glutathione-affinity resins either [14, 15]. Although it has a low affinity for GSH (K_M ~ 0.8 mM), hGST T2-2 is able to activate the tripeptide with an efficiency similar to that of other mammalian GSTs [16, 17]. Moreover, it has been shown that hGST T2-2 presents a sulfatase activity, through GSH conjugation, towards

various carcinogenic sulfate esters and in particular to 1-menaphthyl-sulfate (MSu) [14, 18]. Similar to the hGST T2-2 crystal structure, a sulfate ion has been found bound in the putative active site of domain 1, suggesting the possibility of a sulfatase activity [3, 18]. Considering this information, one could imagine probing domain 1' with a new library based on the MSu scaffold (Figure 3.6. A). For instance, a DCL could be developed based on the commercially available substrate-like 2-naphthaldehyde (Figure 3.6. B).

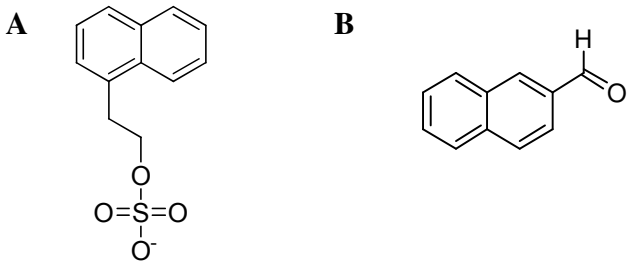


Figure 3.6. Structures of 1-menaphtyl sulfate (**A**) and 2-naphtaldehyde (**B**)

3.3. Conclusions

The N-terminal part of eEF1Bγ from *S. cerevisiae* is a GST-like protein for which no biological function has been assigned so far. Two truncated forms from this region, the monomeric domain 1 and the dimeric domain 1', have been purified using a nickel resin. However, the yields obtained for the two proteins were relatively low and the method needs to be improved. Domain 1' did not influence the CDNB-based DCL composition and is possibly unlikely to bind any CDNB derivative. A potential substrate for this protein was suggested, and a structurally similar commercially available aldehyde has been found to be a good candidate for the generation of a new DCL, which may be able to probe the N-terminal domain of eEF1Bγ.

3.4. Chapter 3 references

1. W.C. Merrick, J.N., *Translational control of gene expression*, ed. J.W.B.H. Nahum Sonenberg, Michael B. Mathews. 2000, New York: Cold Spring Harbor Laboratory Press.
2. Vanwetswinkel, S., et al., *Solution structure of the 162 residue C-terminal domain of human elongation factor 1Bgamma.* J Biol Chem, 2003. **278**(44): p. 43443-51.
3. Jeppesen, M.G., et al., *The crystal structure of the glutathione S-transferase-like domain of elongation factor 1Bgamma from Saccharomyces cerevisiae.* J Biol Chem, 2003. **278**(47): p. 47190-8.
4. Janssen, G.M. and W. Moller, *Elongation factor 1 beta gamma from Artemia. Purification and properties of its subunits.* Eur J Biochem, 1988. **171**(1-2): p. 119-29.
5. Sanders, J., et al., *Immunofluorescence studies of human fibroblasts demonstrate the presence of the complex of elongation factor-1 beta gamma delta in the endoplasmic reticulum.* J Cell Sci, 1996. **109 (Pt 5)**: p. 1113-7.
6. Al-Maghrebi, M., et al., *The 3' untranslated region of human vimentin mRNA interacts with protein complexes containing eEF-1gamma and HAX-1.* Nucleic Acids Res, 2002. **30**(23): p. 5017-28.
7. Kinzy, T.G., T.L. Ripmaster, and J.L. Woolford, Jr., *Multiple genes encode the translation elongation factor EF-1 gamma in Saccharomyces cerevisiae.* Nucleic Acids Res, 1994. **22**(13): p. 2703-7.
8. Ripmaster, T.L., G.P. Vaughn, and J.L. Woolford, Jr., *DRS1 to DRS7, novel genes required for ribosome assembly and function in Saccharomyces cerevisiae.* Mol Cell Biol, 1993. **13**(12): p. 7901-12.
9. Olarewaju, O., et al., *The translation elongation factor eEF1B plays a role in the oxidative stress response pathway.* RNA Biol, 2004. **1**(2): p. 89-94.
10. Hanbauer, I., E.S. Boja, and J. Moskovitz, *A homologue of elongation factor 1 gamma regulates methionine sulfoxide reductase A gene expression in Saccharomyces cerevisiae.* Proc Natl Acad Sci U S A, 2003. **100**(14): p. 8199-204.
11. Koonin, E.V., et al., *Eukaryotic translation elongation factor 1 gamma contains a glutathione transferase domain--study of a diverse, ancient protein superfamily using motif search and structural modeling.* Protein Sci, 1994. **3**(11): p. 2045-54.
12. Kobayashi, S., S. Kidou, and S. Ejiri, *Detection and characterization of glutathione S-transferase activity in rice EF-1betabeta'gamma and EF-1gamma expressed in Escherichia coli.* Biochem Biophys Res Commun, 2001. **288**(3): p. 509-14.
13. Vickers, T.J., S. Wyllie, and A.H. Fairlamb, *Leishmania major elongation factor 1B complex has trypanothione S-transferase and peroxidase activity.* J Biol Chem, 2004. **279**(47): p. 49003-9.
14. Rossjohn, J., et al., *Human theta class glutathione transferase: the crystal structure reveals a sulfate-binding pocket within a buried active site.* Structure, 1998. **6**(3): p. 309-22.
15. Flanagan, J.U., et al., *Mutagenic analysis of conserved arginine residues in and around the novel sulfate binding pocket of the human Theta class glutathione transferase T2-2.* Protein Sci, 1999. **8**(10): p. 2205-12.

16. Hussey, A.J. and J.D. Hayes, *Characterization of a human class-Theta glutathione S-transferase with activity towards 1-menaphthyl sulphate.* Biochem J, 1992. **286 (Pt 3)**: p. 929-35.
17. Caccuri, A.M., et al., *Human glutathione transferase T2-2 discloses some evolutionary strategies for optimization of substrate binding to the active site of glutathione transferases.* J Biol Chem, 2001. **276**(8): p. 5427-31.
18. Jemth, P. and B. Mannervik, *Active site serine promotes stabilization of the reactive glutathione thiolate in rat glutathione transferase T2-2. Evidence against proposed sulfatase activity of the corresponding human enzyme.* J Biol Chem, 2000. **275**(12): p. 8618-24.

Chapter 4: Biotechnological applications of glutathione transferases with quantum dots

4.1. Introduction

4.1.1. Quantum dots and their applications in biology

Quantum dots (QDs) are colloidal nanocrystals of semiconductors with a size range of 1 to 10 nm [1]. They have recently emerged as powerful fluorescent probes for biomolecular and cellular imaging applications. QDs have several advantages over small organic dyes and fluorescent proteins such as size-tuneable photoluminescence, wide excitation range spanning the UV to near infra-red (Figure 4.1.), narrow emission, strong brightness, and high resistance to photobleaching and degradation. QD fluorophores are mainly used in biological imaging, *in vitro* for fixed cells and tissues as well as *in vivo* for live cells and animal imaging [2-4].

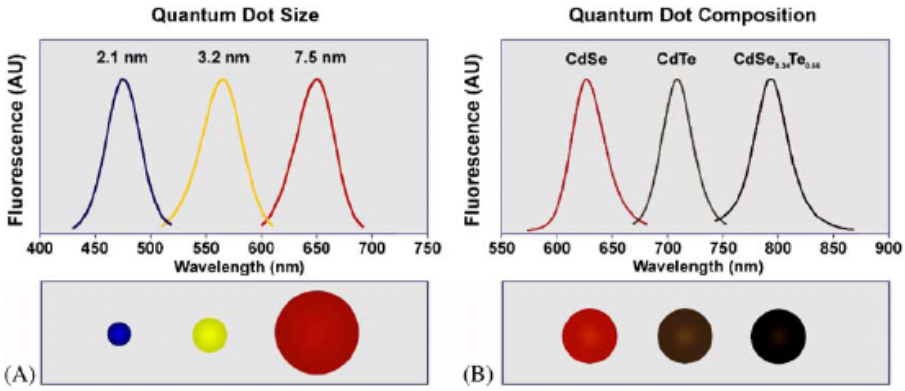


Figure 4.1. Tuning fluorescence emission based on quantum dot size (A) and composition (B) (adapted from Bailey *et al.*; [1])

Cellular labelling with QDs permits extended visualisation of cells under continuous illumination and multicolour imaging. QDs have been used for antigen detection, and labelling of plasma membrane, cytoplasmic and nuclear proteins in fixed cells [5-7]. In live cells, the process is more delicate because cellular viability has to be maintained. QDs are generally taken-up into the live cell by endocytosis, microinjection or electroporation. Targeted-uptake, implying delivery to a specific

cell compartment, is also possible through QDs labelled with specific peptides. Inside live cells, they can be tolerated for extended amounts of time [1, 2]. Using QDs with a long absorption wavelength, one can achieve a deeper penetration of live tissues than the existing dyes. They have been used for mapping, visualising and tracking cancer cells in animals *in vivo* [8, 9].

QD labelling assays have also been developed for detection of DNA, proteins and other biomolecules *in vitro*. DNA-coated QDs have been employed for *in situ* hybridizations [10], probes for human metaphase chromosomes [11], and in single-nucleotide polymorphism and multi-allele DNA detection [12]; while protein-coated QDs have been used in numerous immunoassays [13].

Finally, QDs have been reported as fluorescence resonance energy transfer (FRET) donors. Self-assembling acceptor dye-labelled proteins are attached onto QD donor surfaces, QD donor emission can be size-tuned to improve spectral overlap with a particular acceptor dye. To further improve FRET efficacy, one can use multiple acceptors [14].

QDs for biological applications are generally made of a CdSe core coated over with a protective layer of ZnS. This hydrophobic surface is then modified to make them more water-soluble and to allow attachment of biomolecules. The conjugation of a biomolecule to a QD is generally achieved through a covalent bond, electrostatic or hydrophobic interactions [1, 2]. The ability to control the site of attachment is important to ensure that the biomolecule bound to the QD is still active. Site-specific noncovalent binding of QDs to biomolecules has been achieved and examples include those that exploit carbohydrate-lectin and streptavidin-biotin interactions [15, 16].

4.1.2. Metal-mediated labeling of hexahistidine-tagged proteins

The method to purify proteins with histidine residues was first described in 1987 by Schacher and co-workers [17]. The number of six histidine units was

found to be the most efficient under native conditions and in low or high salt concentrations. Hexahistidine (His_6) tags can generally be incorporated at the termini of proteins with little or no impact on protein function and folding. Since His_6-tags are known to tightly interact with nitrilotriacetic acid (NTA) complexes of transition metal ions such as Cd^{2+}, Hg^{2+}, Co^{2+} and Ni^{2+} (K_d ~ 10 μM), purification of histidine-tagged proteins is accomplished using NTA metal complexes immobilized on resins. After binding, the target protein can be eluted by a gradient of imidazole [18].

Binding affinity between metal-NTA complexes and His_6-tagged proteins was applied to fluorescence labeling of proteins. The first example was reported by Ebright and co-workers who attached the widely used fluorochromes Cy3 and Cy5 to Ni^{2+}:NTA units [19]. Binding of the His_6-tagged transcription factor catabolite activator protein (CAP) to a $(Ni^{2+}:NTA)_2$-Cy3 or $(Ni^{2+}:NTA)_2$-Cy5 resulted in a large increase in fluorescence. The Ni^{2+}:NTA constructs were highly specific for the hexahistidine region. However, these constructs had relatively poor emission, limiting their applications in biological systems. More recently, Lippard and co-workers selectively labeled a His_6-tagged extracellular protein using a fluorescein-NTA conjugate, increasing fluorescence efficiency [20].

4.1.3. SjGST as a tool in biotechnology

GST gene fusion proteins are widely used in proteomics and genomics. Two years after the identification of a M_r 26,000 antigen in mice as a functional GST of the parasitic worm *Schistosoma japonicum* in 1986 [21], Smith and Johnson first described the use of SjGST as a fusion tag for single-step purification of polypeptides [22]. SjGST was cloned in a pGEX *E. coli* expression vector. Bound fusion proteins can be purified from crude lysate by affinity chromatography on immobilized GSH or GSH derivatives, and eluted with GSH. In most cases, the fusion proteins have good aqueous solubility and are therefore

easy to purify. Generally, the fusion proteins are released from the GST-tag by using site specific proteases such as thrombin or the blood coagulation factor Xa, after which the GST carrier and the GST-tagged protease are removed by absorption by GSH affinity chromatography [18].

In the late eighties, this expression system was used to generate recombinant antigens targeting the GST fusion protein. For instance, Rickard and co-workers prepared an antigen useful for vaccinating sheep against ovine cysticercosis [23]. Later, Brunham and co-workers used the same technique in the aim to develop a vaccine against several serovars of *Chlamydia trachomatis*, a bacterium responsible for the chlamydiose disease [24].

More recently, pull-down or affinity precipitation assays using GST fusion proteins have been developed to study protein-protein interactions [25]. A GST fusion protein is expressed, purified and then immobilised on a GSH affinity resin. Once bound, the fusion protein is incubated with cell extracts to pull down the interacting target protein. This technique is widely used for identifying unknown protein-protein interactions or specific regions involved in these interactions. Using this process, Livingstone and co-workers identified a group of proteins interacting with a tumour gene suppressor [26]. In another study, Kaelin and co-workers characterised the sequence-specific DNA binding site domain of the retinoblastoma protein which plays a role in cell growth regulation [27].

4.1.4. Aims and research strategy

The fluorescent labelling of specific proteins using a His_6-tags fusion strategy coupled to the recognition of these tags by Ni-NTA has been recently extended to QDs. During this study, Song and co-workers showed that Ni-NTA-containing QDs can be applied to imaging His_6-tagged proteins in live cells [28]. In a previous study, Mattousi and co-workers showed that His_n-tags can also bind with very high affinity to QDs with carboxylic acid functionalities requiring fewer

synthetic steps and cheaper reagents [29]. However, it was also reported that these QDs bind His_6-tagged proteins only in the presence of Ni^{2+} cations [30]. Therefore, it is useful to look at the advantages, disadvantages and applicability of this fairly new technique. In this study, Manish Gupta and Dr. Juan Mareque-Rivas (University of Edinburgh) investigated binding of the well characterised His_6-tagged and untagged SjGST as model protein to CdSe-ZnS core-shell nanoparticles with carboxylates, NTA, and Ni^{2+}-bound NTA at the surface. For this purpose, I purified and characterised both His_6-tagged and untagged SjGSTs. QDs-GST conjugates are interesting because GSTs catalyze the nucleophilic addition of GSH to the electrophilic center of a range of nonpolar substrates as a way of detoxifying a wide range of harmful endogenous and xenobiotic compounds, and in drug resistance mechanisms. Since GST activity requires not only the formation of a dimeric structure (i.e. protein-protein interactions) but also binding of both GSH and an acceptor substrate (i.e., protein-substrate interactions), it is a good enzyme to investigate the effect of specific and nonspecific binding of QDs on enzymatic activity.

4.2. Results and discussion

4.2.1. Expression, purification and characterisation of SjGST and His_6-SjGST

The untagged SjGST was expressed and purified as described in chapter 1. The N-terminal 6 histidine-tagged-SjGST was expressed at 37 °C in *E. coli*. BL21 (DE3) host cells, using the expression vector pET-6His-SjGST. After induction with IPTG for 3 hours, an increase in expression of the recombinant protein of the expected mass (~28 kDa) was observed by SDS-PAGE analyses (Figure 4.2.).

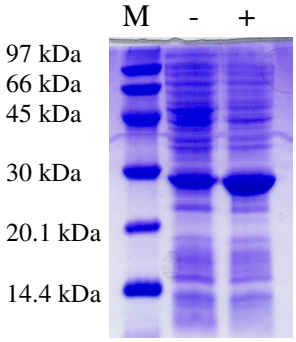


Figure 4.2. SDS-PAGE gel showing expression of His_6-SjGST from *E. coli* before (-) and after (+) induction with IPTG. Clear overexpression of the GST protein (~28 kDa) is observed.

Although pET-6His-SjGST contains a *lacI* gene, a relatively high expression of the recombinant protein is already observed without IPTG. It is likely that the lac I repressor does not bind properly to the lac O operator of the same plasmid, allowing expression without inducer. The His_6-SjGST was purified using a nickel column (immobilized metal affinity chromatography, IMAC), followed by gel filtration chromatography with a 320 ml Sephacryl S-200 gel filtration column. Cell free extracts, affinity columns flow-throughs (containing unbound protein) and washes, and the collected fractions were used for SDS-PAGE analysis (Figure 4.3.).

After the first purification step with a Ni column, a relatively small amount of unbound recombinant protein was found in the flow-through. The bound His_6-SjGST eluted in ten fractions (fractions 19 to 29) which were pooled, concentrated and loaded on a Sephacryl-S200. Three fractions of highly concentrated enzyme were collected from the gel filtration column and a yield of ~ 48 mg was obtained. Despite the two purification steps, two bands of impurities were found in fractions 14 and 15 of the gel filtration column. MALDI-TOF mass spectrometry analysis of the pooled collected fraction was carried out after dialysis against a 20 mM Tris/HCl buffer, containing 1.4 mM of β-mercaptoethanol. The resulting spectrum is presented in Figure 4.4.

Purification of His_6-SjGST

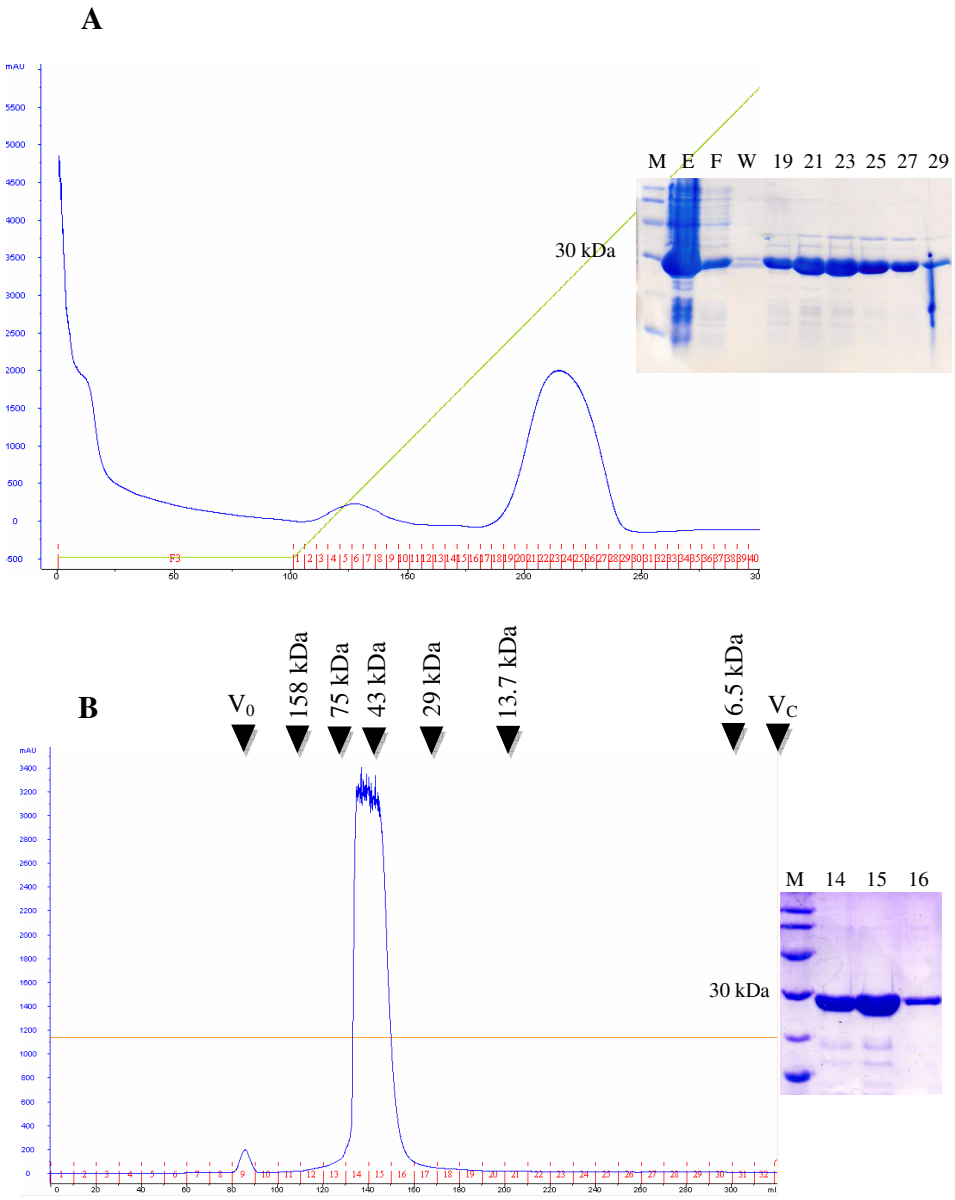


Figure 4.3. Elution profiles of His_6-SjGST from the Ni column (A) and from the Sephacryl-S200 (B) showing the A_{280}. Calibration markers are specified at the top of the S-200 UV trace, V_0 is a void volume, V_C is the column volume. Collected fractions were analysed by SDS-PAGE (M: low molecular weight marker; E: cell free extract; F: flow through; W: wash; collected fractions are represented by their number).

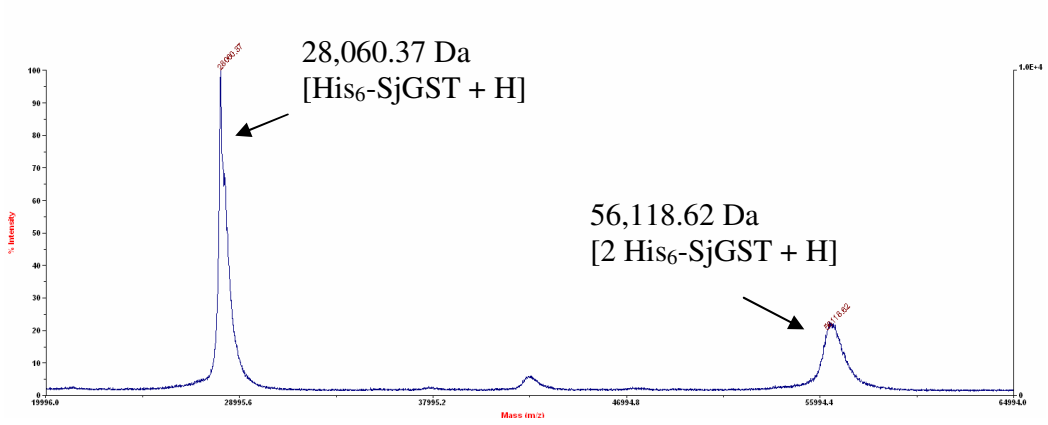


Figure 4.4. MALDI-TOF mass spectrum of His_6-SjGST from 20,000 to 65,000 m/z.

Mass spectrometry analysis revealed that the sample was relatively pure. A first peak showing a mass of 28,060.37 Da was observed which was really close to the theoretical mass of His_6-SjGST plus one proton (28,058 Da). A second peak with a mass of 56,118.62 Da corresponded to the dimeric form of His_6-SjGST plus one proton (56,117 Da).

Kinetic parameters were determined for both the untagged and the His_6-tagged SjGSTs using a CDNB assays. His_6-SjGST data were plotted with Michaelis-Menten and are shown in Figure 4.5. Kinetic parameters towards CDNB for both enzymes are summed up in table 4.1.

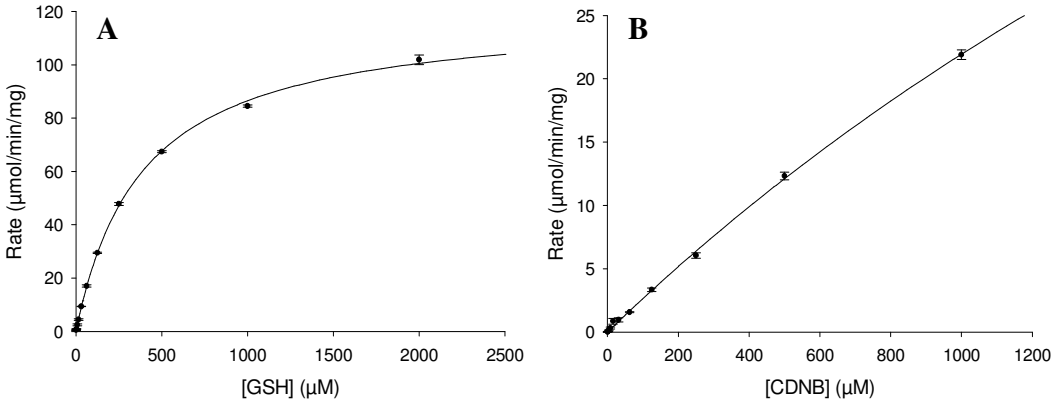


Figure 4.5. Michaelis-Menten representation of His_6-SjGST activity towards the substrates GSH (**A**) and CDNB (**B**). GSH and CDNB concentrations were varied from 1 µM to 2 mM and 0.5 µM to 1 mM, respectively.

Table 4.1. Kinetic parameters for SjGST and His_6-SjGST

	K_M^{GSH} (mM)	K_M^{CDNB} (mM)	k_{cat}^{CDNB} (s^{-1})	k_{cat}/K_M^{CDNB} ($mM^{-1}.s^{-1}$)	Specific Activity ($\mu mol.min^{-1}.mg^{-1}$)
SjGST	0.35 +/- 0.03	4.65 +/- 0.92	50.0 +/- 8.3	10.8	105.1 +/- 17.6
His_6-SjGST	0.39 +/- 0.01	4.32 +/- 0.64	54.6 +/- 7.3	1.3	116.8 +/- 15.6

Kinetic results show that both the untagged and the tagged SjGSTs are active. Moreover, His_6-SjGST displayed similar parameters to the untagged enzyme, meaning that the presence of a histidine tag at the N-terminus of SjGST does not affect its function.

4.2.2. Investigation on His_6-SjGST binding to Ni-NTA capped quantum dots

QDs decorated with Ni-NTA were prepared by Manish Gupta from CdSe-ZnS core-shell dihydrolipoic acid (DHLA)-capped QDs (Figure 4.6.). He used QDs with a maximum fluorescence emission at 614 nm; and found that the

photoluminescence intensity of the Ni-NTA-capped QD is ~ 85% that of the NTA-capped QD. This is an important feature as the photoluminescence is not quenched by the presence of Ni^{2+} in contrast to some organic dyes [19].

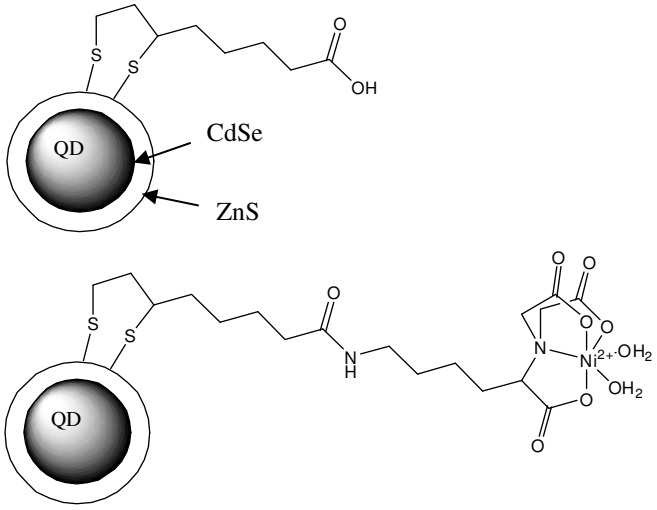


Figure 4.6. Schematic representation of a DHLA-capped QD (top) and a Ni-NTA-capped QD (bottom). Maximum fluorescence emission at 597 nm.

A solution of QDs or PBS buffer (as a control) was incubated with either the untagged or His_6-tagged SjGST for two hours at room temperature. The enzyme-QDs complexes were separated from the unbound enzyme molecules by passing this solution through a 35 nm pore size membrane. The retentate containing the QD-bound enzyme was dissolved in a PBS solution of imidazole to release the enzyme, and both retentate and filtrate were analysed by SDS-PAGE (Figure 4.7.).

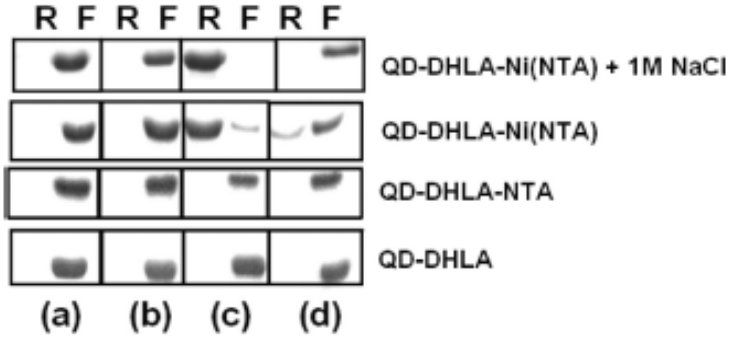


Figure 4.7. SDS-PAGE of the retentate (R) and filtrate (F) after ultrafiltration through a Nanosep 300K filter of His6-GST (a), untagged GST (b), His_6-GST incubated with QD (c), and untagged GST incubated with QD (d). In each case, the enzyme and QD concentrations were 16.5 μM and 9.0 μM, respectively.

In the absence of the QD, His_6-tagged and untagged GST were found only in the filtrate. It has been shown in several studies that DHLA-capped CdSe-ZnS core-shell QDs are able to bind histidine-tagged protein by coordination to Zn^{2+} ions at the nanocrystal surface [29, 31-33]. However, interestingly in this study, no protein was found in the retentate when incubating DHLA-capped QDs with GSTs. Lack of binding could be due to steric hindrance at the N-terminal domain of the histidine tag, preventing access and therefore interaction with the Zn^{2+} of the QD's shell. It is also possible that the different synthetic procedures used to prepare QDs lead to subtle changes at the QD surface which affect binding of biomolecules. It is also important to note that His-tagged proteins can behave differently depending on whether the histidine tag is located at the N- or C-terminus, and that in the studies reporting direct His-tag binding to carboxylate-coated QDs the histidine tag was located at the C-terminus. In contrast, using the same experimental conditions the Ni-NTA-capped QDs immobilized both enzymes and more His_6-tagged than untagged GST was found in the retentate. Binding was also investigated in the presence of high salt concentrations. His_6-GST binding to Ni-NTA-capped QDs was not affected by 1 M NaCl. In contrast, untagged GST did not bind to the QDs under these conditions, which suggests it is predominantly electrostatic. Therefore,

high salt concentrations can be used to avoid binding of untagged proteins while ensuring binding of the desired His_6-tagged target.

The enzyme was easily released from the QD surface upon addition of 0.5 M imidazole, which competes for the Ni^{2+} binding sites. Thus, decorating the surface of the QD with Ni^{2+} complexes of NTA seems a good approach for noncovalent site-specific fluorescent labeling of proteins, which can be used for instance if carboxylate-functionalized QDs lacking Ni^{2+} ions fail. Potential advantages of attaching Ni-NTA units to QDs could be stronger interactions with the His-tag ($Kd \approx 10^{-10}$ M) [34] and less sensitivity to steric hindrance and surface properties by being further away from the nanocrystal surface.

Recently, the value of magnetic nanoparticles as affinity probes to selectively trap and separate His-tagged proteins from cell lysates has been elegantly demonstrated [35-38]. The protein purification efficiency of the Ni-NTA-capped QDs was investigated by incubating cell lysates containing His_6-tagged SjGST for 2 hours. Remarkably, pure fluorescently labelled GST was obtained simply by ultracentrifugaton of this mixture (Figure 4.8.). Thus, by using the Ni-NTA-capped QDs it is possible to purify and fluorescently label His-tagged proteins in a single step. Current methods for efficiently purifying and fluorescently labeling His-tagged proteins need various labour-intensive and expensive steps, such as conjugation of NTA derivatives on support materials or the preparation of suitable magnetic nanoparticles for purification purposes, followed by the attachment of fluorescent tags. Another construct suitable for one-step protein purification and site-specific labelling was recently developed and involves organic fluorophore-doped Ni-NTA-modified silica nanoparticles [39].

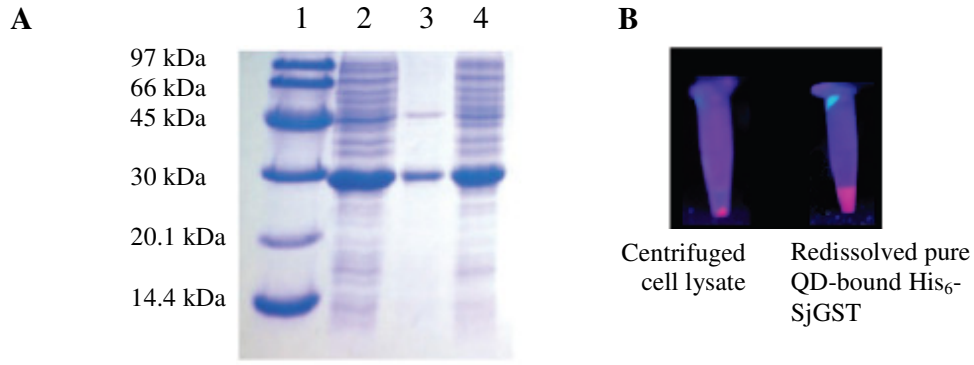


Figure 4.8. A, SDS-PAGE studies of the cell lysate containing His_6-tagged SjGST (lane 2) and proteins released from the Ni-NTA-coated QDs treated with PBS containing 0.5 M imidazole (lane 3) and supernatant (lane 4) after ultracentrifugation. Lane 1 is the molecular weight marker. **B**, Images of the cell lysate after ultracentrifugation and of the pure QD-bound His_6-tagged SjGST.

In order to obtain information about the effect of QD binding on the catalytic activity of GST, Manish Gupta performed GST assays using CDNB. His_6-SjGST and untagged SjGST were incubated with the same concentration of Ni-NTA-capped QD. He found that His_6-tagged GST retained its activity after binding to the QD, whereas the untagged GST lost approximately 22% of its activity (Figure 4.9.). It is suggested that the ability of the His_6-tag to control the position of the Ni-NTA-capped QD relative to the GST active site is responsible for preserving the activity of the enzyme.

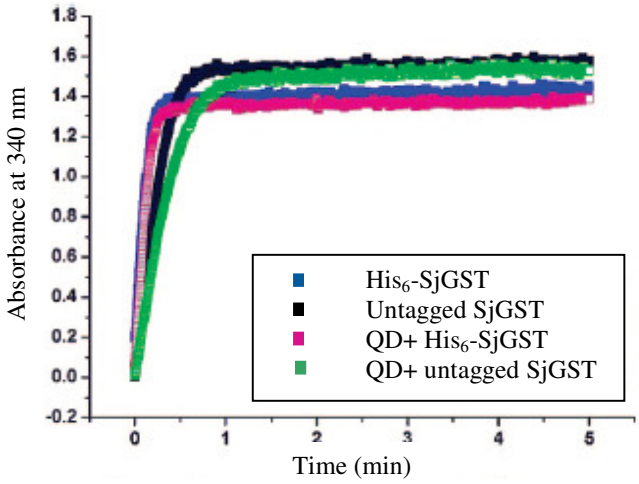


Figure 4.9. Activity of His_6-tagged and untagged SjGST in the absence and in the presence of Ni-NTA coated QDs. The QD alone did not have any activity.

The X-ray crystal structure of SjGST [40] shows that the N-terminus, which is where the His_6-tag was placed, is approximately 25 Å away from the essential catalytic residue Tyr7 (Figure 4.10.). I have examined the distribution of positively and negatively charged residues and found that there are positive and negative regions close to the active site. These are sites where in the absence of the His_6-tag nonspecific electrostatic binding could occur, disrupting the enzyme activity. By comparing the activity of the enzyme which did not bind to the QD with that of the enzyme before incubation with QD, Manish Gupta estimated the protein binding capacity and number of His_6-GST molecules immobilized on each QD (~16). This surface coverage correlates well with that found for QDs and proteins of similar size [32, 33].

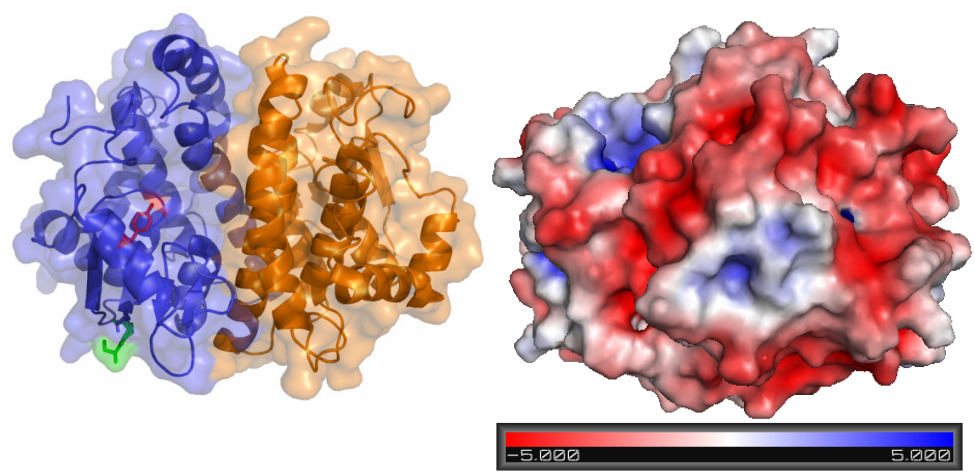


Figure 4.10. X-ray crystal structure of the SjGST homodimer highlighting the catalytically crucial Tyr7 residue in red, and the N-terminus site for the His_6-tag in green (left); surface charge distributions (right). These figures were generated using PyMol.

4.3. Conclusions

Ni-NTA-coated QDs provide a straightforward method to, in one step, purify and reversibly label proteins fluorescently. By using these QDs, the N-terminal His_6-tagged SjGST was selectively labelled and purified, which was not possible using QDs with carboxylates at the surface. Moreover, it was found that Ni^{2+} provides a docking site which helps to precisely orient the fluorescent nanoparticle on the protein surface and that, as a result, GST retained its activity. The use of histidine tags has been broadly adopted in the molecular biology and biochemistry communities, and therefore this specific conjugation strategy should enable widespread use of these QDs for a broad range of biological applications.

4.4. Chapter 4 references

1. Bailey R.E., S.A.M., Nie S., *Quantum dots in biology and medicine.* Physica E, 2004. **25**: p. 1-12.
2. Medintz, I.L., et al., *Quantum dot bioconjugates for imaging, labelling and sensing.* Nat Mater, 2005. **4**(6): p. 435-46.
3. Gao X., Y.L., Petros J.A., Marshall F.F., Simons J.W. and Nie S., *In vivo molecular and cellular imaging with quantum dots.* Current Opinion in Biotechnology, 2005. **16**: p. 63-72.
4. Michalet, X., et al., *Quantum dots for live cells, in vivo imaging, and diagnostics.* Science, 2005. **307**(5709): p. 538-44.
5. Bruchez, M., Jr., et al., *Semiconductor nanocrystals as fluorescent biological labels.* Science, 1998. **281**(5385): p. 2013-6.
6. Chan, W.C. and S. Nie, *Quantum dot bioconjugates for ultrasensitive nonisotopic detection.* Science, 1998. **281**(5385): p. 2016-8.
7. Chen, L., et al., *Quantum-dots-based fluoroimmunoassay for the rapid and sensitive detection of avian influenza virus subtype H5N1.* Luminescence, 2009.
8. Kim, S., et al., *Near-infrared fluorescent type II quantum dots for sentinel lymph node mapping.* Nat Biotechnol, 2004. **22**(1): p. 93-7.
9. Gao, X., et al., *In vivo cancer targeting and imaging with semiconductor quantum dots.* Nat Biotechnol, 2004. **22**(8): p. 969-76.
10. Pathak, S., et al., *Hydroxylated quantum dots as luminescent probes for in situ hybridization.* J Am Chem Soc, 2001. **123**(17): p. 4103-4.
11. Xiao, Y. and P.E. Barker, *Semiconductor nanocrystal probes for human metaphase chromosomes.* Nucleic Acids Res, 2004. **32**(3): p. e28.
12. Gerion, D., et al., *Room-temperature single-nucleotide polymorphism and multiallele DNA detection using fluorescent nanocrystals and microarrays.* Anal Chem, 2003. **75**(18): p. 4766-72.
13. Goldman, E.R., et al., *Avidin: a natural bridge for quantum dot-antibody conjugates.* J Am Chem Soc, 2002. **124**(22): p. 6378-82.
14. Medintz, I.L., et al., *A fluorescence resonance energy transfer-derived structure of a quantum dot-protein bioconjugate nanoassembly.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9612-7.
15. Babu, P., S. Sinha, and A. Surolia, *Sugar-quantum dot conjugates for a selective and sensitive detection of lectins.* Bioconjug Chem, 2007. **18**(1): p. 146-51.
16. Howarth, M., et al., *Targeting quantum dots to surface proteins in living cells with biotin ligase.* Proc Natl Acad Sci U S A, 2005. **102**(21): p. 7583-8.
17. Hochuli E, B.W., Döbeli H, Gentz R, Stüber D, *Genetic approach to facilitate purification of recombinant*
proteins with a novel metal chelate adsorbent. Bio/Technology, 1988. **6**: p. 1321-1325.
18. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2003. **60**(5): p. 523-33.

19. Kapanidis, A.N., Y.W. Ebright, and R.H. Ebright, *Site-specific incorporation of fluorescent probes into protein: hexahistidine-tag-mediated fluorescent labeling with (Ni(2+):nitrilotriacetic Acid (n)-fluorochrome conjugates.* J Am Chem Soc, 2001. **123**(48): p. 12123-5.
20. Goldsmith, C.R., et al., *Selective labeling of extracellular proteins containing polyhistidine sequences by a fluorescein-nitrilotriacetic acid conjugate.* J Am Chem Soc, 2006. **128**(2): p. 418-9.
21. Smith, D.B., et al., *Mr 26,000 antigen of Schistosoma japonicum recognized by resistant WEHI 129/J mice is a parasite glutathione S-transferase.* Proc Natl Acad Sci U S A, 1986. **83**(22): p. 8703-7.
22. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene, 1988. **67**(1): p. 31-40.
23. Johnson, K.S., et al., *Vaccination against ovine cysticercosis using a defined recombinant antigen.* Nature, 1989. **338**(6216): p. 585-7.
24. Toye, B., et al., *Immunologic characterization of a cloned fragment containing the species-specific epitope from the major outer membrane protein of Chlamydia trachomatis.* Infect Immun, 1990. **58**(12): p. 3909-13.
25. Ren, L., et al., *Glutathione S-transferase pull-down assays using dehydrated immobilized glutathione resin.* Anal Biochem, 2003. **322**(2): p. 164-9.
26. Kaelin, W.G., Jr., et al., *Identification of cellular proteins that can interact specifically with the T/E1A-binding region of the retinoblastoma gene product.* Cell, 1991. **64**(3): p. 521-32.
27. Chittenden, T., D.M. Livingston, and W.G. Kaelin, Jr., *The T/E1A-binding domain of the retinoblastoma product can interact selectively with a sequence-specific DNA-binding protein.* Cell, 1991. **65**(6): p. 1073-82.
28. Kim, J., et al., *Ni-nitrilotriacetic acid-modified quantum dots as a site-specific labeling agent of histidine-tagged proteins in live cells.* Chem Commun (Camb), 2008(16): p. 1910-2.
29. Sapsford, K.S., Pons, T., Medintz, I. L., Higashiya, S., Brunel, F. M., Dawson, P. E., and Mattoussi, H., *Kinetics of metal-affinity driven self-assembly between proteins or peptides and CdSe-ZnS Quantum dots.* Journal of Physical Chemistry C, 2007. **111**: p. 11528-11538.
30. Yao, H., et al., *Quantum dot/bioluminescence resonance energy transfer based highly sensitive detection of proteases.* Angew Chem Int Ed Engl, 2007. **46**(23): p. 4346-9.
31. Ipe, B.I. and C.M. Niemeyer, *Nanohybrids composed of quantum dots and cytochrome P450 as photocatalysts.* Angew Chem Int Ed Engl, 2006. **45**(3): p. 504-7.
32. Mattoussi, H., Mauro, J. M., Goldman, E. R., Anderson, G. P., Sundar, V. C., Mikulec, F. V., and Bawendi, M. G. , *Self-Assembly of CdSe-ZnS Quantum Dot Bioconjugates Using an Engineered Recombinant Protein.* Journal of the American Chemical Society, 2000. **122**: p. 12142-12150.
33. Medintz, I.L., et al., *Self-assembled nanoscale biosensors based on quantum dot FRET donors.* Nat Mater, 2003. **2**(9): p. 630-8.
34. Hainfeld, J.F., et al., *Ni-NTA-gold clusters target His-tagged proteins.* J Struct Biol, 1999. **127**(2): p. 185-98.

35. Lee, I.S., et al., *Ni/NiO core/shell nanoparticles for selective binding and magnetic separation of histidine-tagged proteins.* J Am Chem Soc, 2006. **128**(33): p. 10658-9.
36. Xu, C., et al., *Nitrilotriacetic acid-modified magnetic nanoparticles as a general agent to bind histidine-tagged proteins.* J Am Chem Soc, 2004. **126**(11): p. 3392-3.
37. Lee, K.B., S. Park, and C.A. Mirkin, *Multicomponent magnetic nanorods for biomolecular separations.* Angew Chem Int Ed Engl, 2004. **43**(23): p. 3048-50.
38. Lee, K.S. and I.S. Lee, *Decoration of superparamagnetic iron oxide nanoparticles with Ni2+: agent to bind and separate histidine-tagged proteins.* Chem Commun (Camb), 2008(6): p. 709-11.
39. Kim, S.H., M. Jeyakumar, and J.A. Katzenellenbogen, *Dual-mode fluorophore-doped nickel nitrilotriacetic acid-modified silica nanoparticles combine histidine-tagged protein purification with site-specific fluorophore labeling.* J Am Chem Soc, 2007. **129**(43): p. 13254-64.
40. McTigue, M.A., D.R. Williams, and J.A. Tainer, *Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica and its complex with the leading antischistosomal drug praziquantel.* J Mol Biol, 1995. **246**(1): p. 21-7.

Chapter 5: Materials and methods

5.1. General Materials

5.1.1. General reagents

All reagents, chemicals and media were purchased from Sigma, Aldrich, Fisher, Biorad, Pharmacia or Oxoid unless otherwise stated. All competent cells and plasmids were purchased from Novagen and all chromatography columns were from GE Healthcare unless otherwise indicated. The pET-6His-SjGST plasmid was constructed in the laboratory, prior to this work, by Dr. Dominic Campopiano. Plasmids pET15-b-mGSTM1 and pET9-a-mGSTA4 were generously supplied by Prof. John Hayes (University of Dundee, UK). Plasmid pET15-b-hGSTP1 was a kind gift from Dr. Sylvie Blond (University of Illinois, USA). Pasmids pET11-d-TKB587 and pET11-d-TKB611 were provided by Dr. Gregers Andersen (University of Arhus, Denmark).

5.1.2. Media and Solutions

Sterilisation of media: all media were autoclaved at 121°C for 20 minutes prior to use.

Luria Bertani (LB): tryptone (10 $g.l^{-1}$), yeast extract (5 $g.l^{-1}$), sodium chloride (10 $g.l^{-1}$); pH adjusted to 7.5 with sodium hydroxide.

2 x YT: tryptone (16 $g.l^{-1}$), yeast extract (10 $g.l^{-1}$), sodium chloride (5 $g.l^{-1}$); pH adjusted to 7.5 with sodium hydroxide.

SOC: tryptone (20 $g.l^{-1}$), yeast extract (5 $g.l^{-1}$), sodium chloride (0.5 $g.l^{-1}$), magnesium sulfate (5 $g.l^{-1}$), glucose (3.2 $g.l^{-1}$); pH adjusted to 7.5 with sodium hydroxide.

Agar plates: LB Agar (35 $g.l^{-1}$) was dissolved in the appropriate volume of deionised water. The specific antibiotic was added prior to making plates.

X-Gal plates: S-Gal™/LB Agar Blend from Sigma-Aldrich containing tryptone (10 $g.l^{-1}$), yeast extract (5 $g.l^{-1}$), sodium chloride (10 $g.l^{-1}$), agar (12 $g.l^{-1}$), S-Gal (0.3 $g.l^{-1}$), ferric ammonium citrate (0.5 $g.l^{-1}$), IPTG (0.03 $g.l^{-1}$) was dissolved in the appropriate amount of deionised water according to the manufacturer's instructions. The specific antibiotic was added prior to making plates.

DNA-running buffer: TAE buffer – 40 mM Tris, 20 mM Acetic acid, 1 mM EDTA (pH 8.3).

Protein-running buffer: TGS buffer – 25 mM Tris, 192 mM glycine, 0.1% (w / v) SDS (pH 8.3).

5.1.3. Purification buffers

Purification of His_6-hGSTP1-1, His_6-mGSTM1-1, His_6-SjGST, His_6-TKB587 and His_6-TKB611

- Buffer A (binding buffer): 20 mM Tris-HCl, 0.5 M NaCl, 5 mM imidazole (pH 6.8)
- Buffer B (elution buffer): 20 mM Tris- HCl, 0.5 M NaCl, 0.5 M imidazole (pH 6.8)
- Buffer C (gel filtration buffer) : 0.1 M KPhos buffer (0.0503 M KH_2PO_4, 0.0497 M K_2HPO_4), 150 mM NaCl (pH 6.8)
- Buffer D (dialysis buffer): 0.1 M KPhos buffer (pH 6.8)

Purification of SjGST, SjGST-Y7F, mGSTA4-4 and BphK_J2315

- Buffer E (binding buffer): 20 mM Tris-HCl (pH 7)
- Buffer F (elution buffer): 20 mM Tris-HCl, 10 mM glutathione (pH 8)
- Buffer C (gel filtration buffer) : 0.1 M KPhos buffer (0.0503 M KH_2PO_4, 0.0497 M K_2HPO_4), 150 mM NaCl (pH 6.8)
- Buffer D (dialysis buffer): 0.1 M KPhos buffer (pH 6.8)

5.2. Molecular Biology

5.2.1. Bacterial cell lines

Strain	**Genotype**	**Applications**
Top 10™ (Competent)	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L (Str^R) *nup*G	Cloning
DH5α™ (Competent)	F⁻ ϕ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *deo*R *rec*A1 *end*A1 *hsd*R17 (r_{k-}, m_{k+}) *pho*A *sup*E44Λ⁻ *thi*-1 *gyr*A96 *rel*A1	DNA isolation
JM109 (Competent)	*end*A1 *rec*A1 *gyr*A96 *thi hsd*R17 (r_{k-}, m_{k+}) *rel*A1 *sup*E44, Δ(*lac-pro*AB), [F' *tra*D36 *pro*AB, *laq*IqZΔM15]	Protein expression Transformation of DNA ligations
BL21(DE3) (Competent)	F⁻ *omp*T hsdS_B (r_B^- m_B^-) *gal dcm* (DE3)	Protein expression

5.2.2. Plasmids pGEX6P-1, pET-6His-SjGST, pET9-a-mGSTA4, pET15-b-mGSTM1, pET15-b-hGSTP1, pET11-d-TKB588 and pET11-d-TKB611

The plasmid pGEX-6P-1 used for expression of SjGST was obtained from Amersham Biosciences (Genbank Accession number U78872; figure 5.1.).

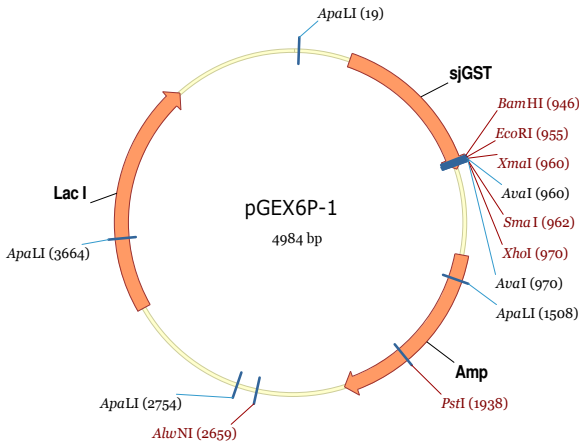


Figure 5.1. The plasmid pGEX6P-1 used for expression of SjGST.

Plasmid pET-6His-SjGST used for expression of the N-terminal 6-histidine tagged SjGST was generated previously from the modification of pGEX6P-1, in our laboratory (Figure 5.2.).

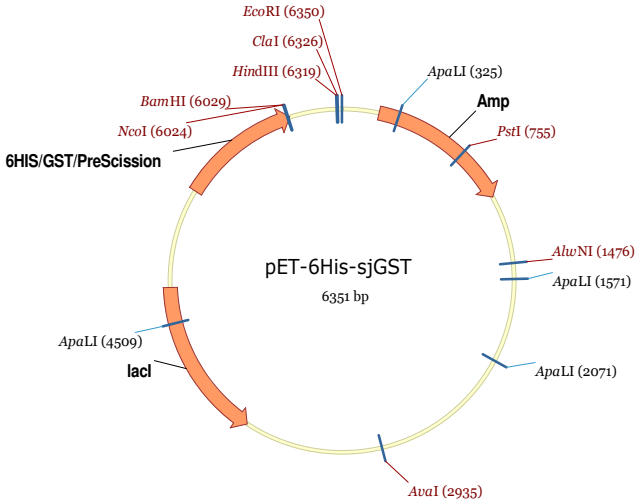


Figure 5.2. The plasmid pET-6His-SjGST used for expression of His_6-SjGST.

Plasmids pET9-a-mGSTA4 and pET15-b-mGSTM1 were kindly provided by Prof. John Hayes and his research group based in the University of Dundee (Figures 5.3. and 5.4.).

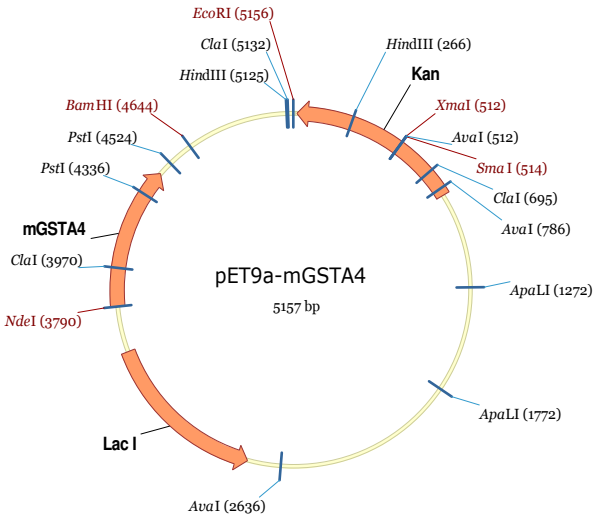


Figure 5.3. The plasmid pET9-a-mGSTA4 used for expression of mGST A4-4.

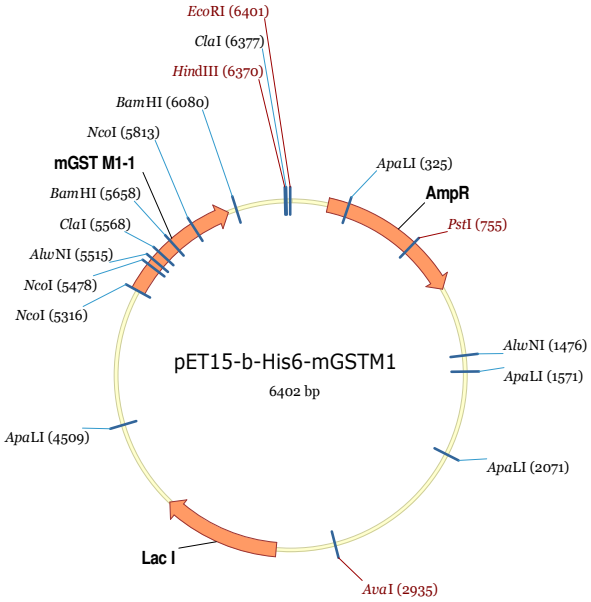


Figure 5.4. The plasmid pET15-b-His_6-mGSTM1 used for expression of His_6-mGST M1-1.

Plasmid pET15-b-His_6-hGSTP1 was kindly provided by Dr. Sylvie Blond and his research group based in the University of Illinois, Chicago, USA (Figure 5.5.) [1].

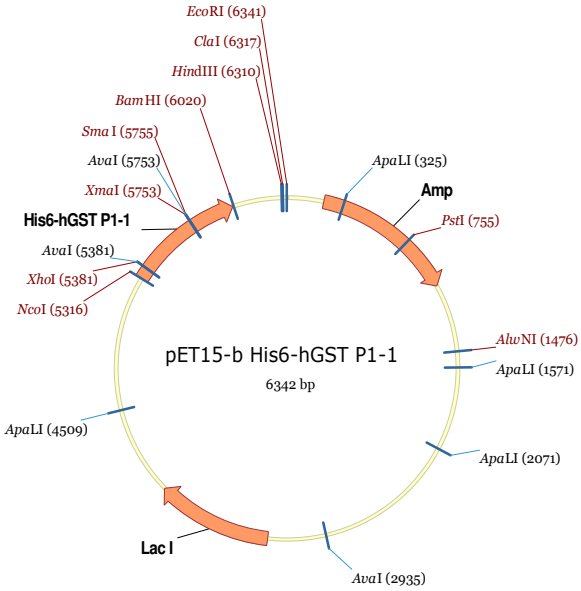


Figure 5.5. The plasmid pET15-b-His_6-hGSTP1 used for expression of His_6-hGST P1-1.

Plasmids pET11-d-TKB588 and pET11-d-TKB611 were generously provided by Dr. Gregers Andersen and his research group based in the University of Aahrus, Denmark (Figures 5.6. and 5.7.) [2].

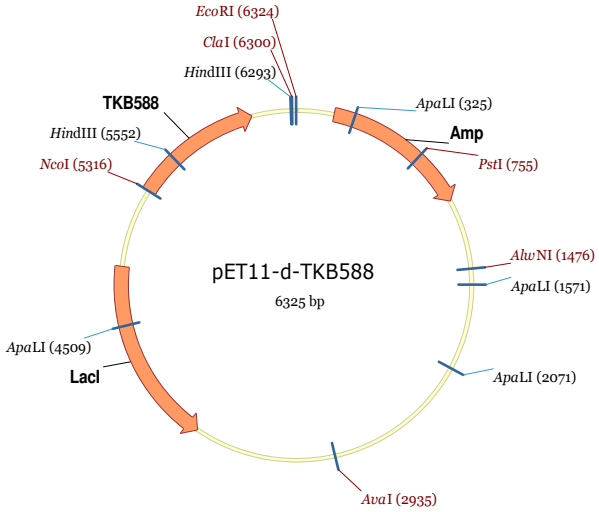


Figure 5.6. The plasmid pET11-d-TKB588 used for expression of the 588 N terminus truncated eEF1Bγ from *Saccharomyces cereviseae.*

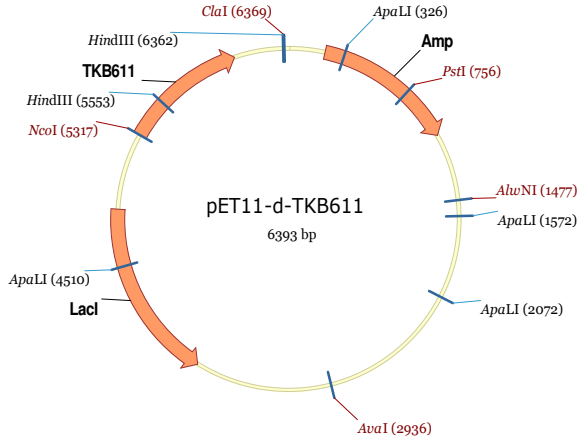


Figure 5.7. The plasmid pET11-d-TKB611 used for expression of the 611 N terminus truncated eEf1bγ from *Saccharomyces cereviseae.*

5.2.3. Oligonucleotide primers

The following oligonuclotide primers were used in this study. Restriction sites are underlined and mutation sites are in bold.

Name	Sequence (5'-3')
BphK_J2315 Forward	GAC ACC CAC GAA GGA ACC ATA TGA AGC T
BphK_J2315 Reverse	GCA CCG CTG CGG CGG CCA CGC GTC ATC G
BphK_J2315_NcoI Forward	CCT GTT GCG ATG ACC CAT GGC CGC CGC A
BphK_J2315_NcoI Reverse	TGC GGC GGC CAT GGG TCA TCG CAA CAG G
pGEX Forward	GGG CTG GCA AGC CAC GTT TGG TG
pGEX Reverse	CCG GGA GCT GCA TGT GTC AGA GG
pUC/M13 Forward	GTT TTC CCA GTC ACG AC
pUC/M13 Reverse	CAG GAA ACA GCT ATG AC
SjGST_Y7F Forward	CCT ATA CTA GGT TTT TGG AAA ATT AAG
SjGST_Y7F Reverse	CTT AAT TTT CCA AAA ACC TAG TAT AGG
T7 Promoter	TAA TAC GAC TCA CTA TAG
T7 Terminator	CTA GTT ATT GCT CAG CGG

5.2.4. DNA manipulation

5.2.4.1. Purification of plasmid DNA

Plasmid DNA was prepared using QIAprep® Spin Miniprep Kit (Qiagen) following the manufacturer's instructions, provided with the kit and summarized below.

An over-night culture (1.5 mL) was transferred to a sterile epindorf (1.5 mL) and centrifuged (13,000 rpm, 5 minutes) using a bench-top centrifuge. The supernatant was discarded and the procedure repeated with a further 1.5 mL of culture. The pellet was resuspended in P1 buffer (250 µL) and P2 buffer (250 µL) was added. This was

followed by immediate invertion of the epindorf 4-6 times. Subsequently, N3 buffer (350 µl) was added and the mixture inverted immediately (4-6 times). The resultant cloudy mixture was centrifuged (13,000 rpm, 10 minutes) after which, the supernatant was transferred to a QIAprep spin column and centrifuged (13,000 rpm, 1 minute). The flow-through was discarded and the spin column washed with PB buffer (500 µL) and centrifuged (13,000 rpm, 1 minute). The column was washed further with PB buffer (750 µL) and centrifuged (13,000 rpm, 1 minute). The column was centrifuged for one more minute to remove any residual wash buffer before the purified DNA was eluted with sterile water (100 µL) and stored at -20˚C until required.

5.2.4.2. Tranformation of *E. coli* competent cells with recombinant DNA

Competent cells were transformed according to the manufacturer's instructions. DNA (up to 40 ng) was added to an aliquot of competent cells and gently mixed. This was left on ice for 2 minutes before the cells were heat shocked (42°C, 30 sec). The cells were then grown in 80 µl SOC medium at 37°C for 1 hr. Finally, the cells were spread to dryness on selective agar plates and incubated at 37°C overnight.

5.2.4.3. Electrophoresis of DNA

The required amount of agarose was added to TAE buffer (typically 1 g / 100 ml) and heated at 100°C until dissolved. The solution was allowed to cool to 55°C and ethidium bromide was added to a final concentration of 0.5 $\mu g.ml^{-1}$. The gel was then poured into a casting mould and allowed to set at room temperature. Loading dye (Promega) was added to the sample prior to loading the DNA migrated using a potential difference of 100 V for an adequate time to achieve separation. The gels were viewed and photographed under UV-light. HyperLadder I and IV (Bioline) were used as molecular weight markers.

5.2.4.4. Digestion of DNA with restriction endonucleases

The required amount of DNA (0.5 – 1 μg) was treated with the appropriate amount of endonuclease and buffer, and incubated for at least 3 hours at 37°C before analysis by electrophoresis on agarose. The restriction enzyme *EcoR*I was used as a control for the digestion of pGEM-T easy plasmids.

5.2.4.5. Gel-extraction of DNA

DNA was purified from agarose using QIAquick® Gel Extraction Kit (Qiagen) following the manufacturers' instructions which are summerised below.

Agarose, containing the DNA, was covered of QG buffer (600 μL – 1 mL) and incubated at 50 ˚C until the agarose had completely dissolved. This solution was transferred to a spin column and centrifuged (13,000 rpm, 1 minute) and the supernatant discarded. QG buffer (500 μL) was added and the column centrifuged again (13,000 rpm, 1 minute). The supernatant was discarded and PE buffer (750 μL) added to the column and incubated for 2-5 minutes. The column was centrifuged (13,000 rpm, 1 minute), the supernatant discarded, and the column centrifuged again to remove any residual buffer. The purified DNA was eluted with sterile water (30 μL) and stored at -20˚C until required.

5.2.4.6. Direct cloning of PCR products

All PCR products were cloned into the pGEM-T easy vector (Promega) using the manufacturers' instructions. 2x rapid T4 DNA ligase buffer (2 μl), pGEM-T easy vector (1 μl), PCR product (3 μl) and T4 DNA ligase from Promega (1 μl) were gently mixed and incubated for one hour at room temperature. 2 μl of the reaction mixture was used to transform JM109 competent cells as described in section 5.2.4.2.

5.2.4.7. Cloning into plasmid vectors

The DNA fragment cut with suitable restriction enzymes (8 µl), the host vector cut with suitable restriction enzymes (2 µl), 2x Quick ligation buffer (2 µl) and Quick T4 DNA ligase from New England Biolabs (1 µl) were gently mixed and incubated for 10 min at room temperature. The reaction mixture (3 µl) was used to transform JM109 competent cells as described in section 5.2.4.2.

5.2.4.8. Storage of bacterial stocks

LB medium (5 mL) containing the appropriate antibiotic was inoculated with the strain of interest and incubated over-night at 37 °C. Subsequently, 225 µL of sterile 80% glycerol and 1 mL of bacterial culture were placed into a labeled 1.5 mL cryotube, giving a final bacterial stock containing 15% glycerol. The tube was quickly vortexed and placed at -80°C until required.

5.2.4.9. Polymerase chain reactions

Polymerase chain reactions (PCRs) were performed using a Techne TC-3000 thermal cycler.

5.2.4.9.1. Amplification of DNA

A typical amplification PCR contained two Ready to Go PCR™ beads (Amersham Biosciences), DNA template (2 µl), primer-forward 10 µM (5 µl), primer-reverse 10 µM (5 µl), and distilled water (final volume of 50 µl). The cycling parameters of the PCR are listed below:

PCR steps	Temperature (°C)	Time (min)	Number of cycles
Initial denaturation	95	2	1
Denaturation	95	1	25
Annealing	55	1	25
Extension	72	2	25
Termination	72	10	1

The PCR product was then subjected to agarose gel electrophoresis and the required band was excised. DNA was purified as described in section 5.2.4.5.

5.2.4.9.2. DNA sequencing

The sequencing reactions were performed using the BigDye® Terminator v3.1 Cycle sequencing kit (PE Applied Biosystems). The PCR mixtures contained DNA template ~5 pmol (5 µl), 5x reaction buffer (2 µl), primer 10 µM (1 µl) and Big Dye 3.1 (2 µl). The pET, pUC/M13 and pGEX primers were used to sequence the pET, pGEM and pGEX inserts, respectively. Sequencing of the 5'-end of the DNA template was carried out with the primer forward while the sequencing of the 3'-end was carried out with the primer reverse. The cycling parameters of the PCR are listed below.

PCR steps	Temperature (°C)	Time	Number of cycles
Denaturation	95	30 sec	16
Annealing	60	20 sec	16
Extension	68	4 min	16

Automated DNA sequencing was performed on an ABI prism 377 DNA sequencer using the Sanger dideoxy chain termination method. Sequence data were analysed using Contig Express within Vector NTI Advance™ V10 software package.

5.2.4.9.3. Site-directed mutagenesis

Mutations were performed by using Stratagene site-directed mutagnenesis kit. Each reaction contained plasmid DNA (5 µl), 10x reaction buffer (5 µl), primer-forward (1 µl), primer-reverse (1 µl), dNTP mix (1 µl), distilled water (final volume 50 µl) and Pfu DNA polymerase 2.5 u / µl (1 µl). The amplification parameters are listed below. One microliter (5 U) of *Dpn*I restriction enzyme was added to the PCR product and incubated at 37°C for 2 h. The reaction products were used to transform competent *E. coli* cells DH5α (DE3).

PCR steps	**Temperature (°C)**	**Time**	**Number of cycles**
Initial denaturation	95	2 min	1
Denaturation	95	30 sec	16
Annealing	55	60 sec	16
Extension	68	1 min / 1000 bp	16
Termination	68	7 min	1

5.2.4.9.4. Generation of a SjGST_Y7F mutant

A site-directed mutagenesis was carried out as described in section 5.2.4.9.3, using plasmid pGEX-6P-1 as a template. SjGST_Y7F Forward and Reverse primers were used to change the TAT codon in the SjGST sequence to TTT, so that the

tyrosine amino-acid residue was replaced by a phenylalanine (SjGST_Y7F). Transformed DH5α were spread on agar plates containing ampicillin (100 µg/mL). Resulting colonies were used to inoculate 5 mL LB liquid media with ampicillin (100 µg/mL) and cultured over-night at 37 ˚C. Subsequently, plasmid vectors were purified as described in section 5.2.4.1. and sequenced as described in section 5.2.4.9.2. pGEX6P-1 vectors containing the mutation were named pGEX-6P-1-Y7F (Figure 5.8.) and kept at -20˚C until required.

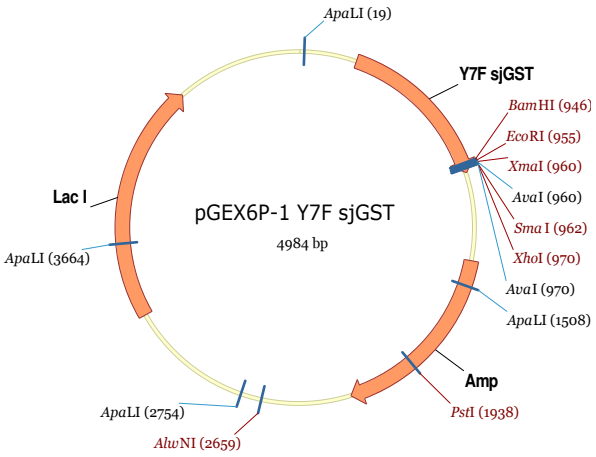


Figure 5.8. The plasmid pGEX6P-1-Y7F used for expression of SjGST_Y7F

5.2.4.9.5. Isolation and cloning of *bphK_J2315* in pET22-b

Isolation of *bphK_J2315* from *Burkholderia cenocepacia* genome

B. cenocepacia J2315 genomic DNA was provided by Dr. Josefin Bartholdson from our laboratory. The known BphK protein sequence from *B. xenovorans* strain LB400 was used to search the genome of *B. cenocepacia* for similar translated sequences. The highest match was obtained with the gene BCAM0431 sharing 45% identity. The BphK_J2315 Forward and Reverse primers were designed

subsequently and used to amplify *bphK_J2315* from the J2315 genome as described in section 5.2.4.9.1.

The PCR product was ligated into pGEM®-T Easy vector according to the manufacturer's instructions summarized below. The PCR product (2 µL, ~25 ng) was added to 5 µL of 2X rapid ligation buffer, 1 µL of pGEM®-T Easy plasmid, 1 µL of T4 DNA ligase and 11 µL of distilled and filtered water. The reaction was mixed by pipetting and incubated for 1 h at room temperature.

After incubation, 2 µl of reaction were used to transfom JM109 competent cells as described in section 5.2.4.2. Transformed cells were spread on X-Gal plates containing ampicillin (100 µg/mL) for a subsequent blue:white screening. After over-night incubation at 37˚C, several white colonies were used to inoculate 5 mL LB liquid media with ampicillin (100 µg/mL) and cultured over-night at 37 ˚C. Subsequently, plasmid vectors were isolated as described in section 5.2.4.1. and sequenced as described in section 5.2.4.9.2. pGEM vectors containing *bphk_J2315* were kept at -20˚C until required.

Cloning of *bphk_J2315* in pET22-b

In order to insert the gene in a pET22-b expression vector, a site-directed mutagenesis was carried out to introduce an NcoI restriction site at its C-terminus, using BphK_J2315_NcoI Forward and Reverse primers as described in section 5.2.4.9.3. The PCR product was used to transform DH5α (DE3) competent cells which were subsequently spread on agar plates containing ampicillin (100 µg/mL). After over-night incubation at 37 ˚C, several colonies were picked and used to inoculate 5 mL of LB medium with ampicillin (100 µg/mL) each. Plasmids were isolated with a QIAprep® Spin Miniprep Kit as described in section 5.2.4.1. and subjected to digestion using NdeI and NcoI restriction enzymes as described in section 5.2.4.4. Restriction products were separated by gel electrophoresis. DNA fragments

with a size corresponding to that of *bphk_J2315* (624 bp) were extracted from the agarose gel as described in section 5.2.4.5. and cloned in pET22-b as described in section 5.2.4.7. The insertion was confirmed by restriction analyses using NdeI and NcoI restriction enzymes and DNA sequencing. The resulting vector containing the gene was named pET22-b-BphK_J2315 (Figure 5.9.) and was kept at -20˚C until required.

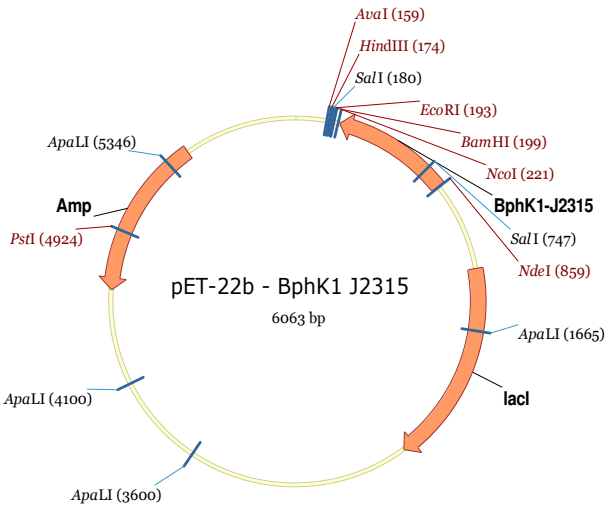


Figure 5.9. The plasmid pET22-b-BphK_J2315 used for expression of BphK_J2315

5.3. Protein work

5.3.1. Polyacrylamide Gel Electrophoresis (PAGE)

SDS-PAGE was used to analyse proteins on the basis of their molecular mass with a Tris-Glycine or "Laemmli" discontinuous buffering system [3].

SDS sample 2x loading buffer contained Tris/HCl (1.5 M, pH 8.0, 1 ml), glycerol (2 ml), bromophenol blue (0.05%, 2 ml), SDS (10%, 1.6 ml) and β-mercaptoethanol (0.4 ml). Samples for analysis by SDS-PAGE were prepared by

addition of appropriate volume of this buffer, followed by heating (100°C, 5 min) and centrifugation (13,000 rpm, 5 min).

A standard method was used for the production of acrylamide gels with a running gel containing 15% acrylamide and a stacking gel of 4% acrylamide. Briefly, the running gel (acrylamide/bis (29:1) 15% w/v, SDS 0.1 % w/v, TEMED 0.15% v/v and APS 0.1% w/v in Tris buffer (375 mM, pH 8.8)) and was poured between glass plates, levelled and set at room temperature. The stacking gel (acrylamide/bis (29:1) 4% w/v, SDS 0.1 % w/v, TEMED 0.15 % v/v and APS 0.1% w/v in Tris buffer (375 mM, pH 6.8)) was then added and set at room temperature using a mould to produce the wells in the finished gel. Gels were run in protein-running buffer at 150 V and then visualised using Coomassie Brilliant Blue R250, GelCode (Pierce) in accordance with manufacturers' instructions.

5.3.2. Expression and purification of glutathione transferases

5.3.2.1. Large scale expression

A general method, as follows, was used to express the GST isoforms and the GST-like proteins.

The appropriate plasmid vector was used to transform *E. coli* expression strain BL21 (DE3). One colony was picked and used to inoculate 200 ml of LB broth containing ampicillin (100 μg/ml) or kanamycin (30 μg/ml) and grown overnight at 37°C with agitation (250 rpm). The 200 ml overnight culture was used to inoculate 3/5 L of appropriate media, grown to an OD_{600} of 0.6 at 37°C and induced with IPTG for three to five hours at the appropriate temperature. The conditions of expression for each protein are detailed below:

Enzyme	Plasmid	Antibiotic resistance	Culture media	[IPTG] required (mM)	Time of growth after induction	Temperature of growth after induction
SjGST	pGEX-6-P1	Amp	2YT	0.1	3-4 h	37 °C
His_6-SjGST	pET-6His-SjGST	Amp	LB broth	0.1	3-4 h	37 °C
SjGST_Y7F	pGEX-6-P1-Y7F	Amp	2YT	0.1	5-6 h	30 °C
His_6-mGSTM1-1	pET15-b-His_6-mGSTM1	Amp	2YT	1.0	3-4 h	37 °C
mGSTA4-4	pET9-a-mGSTA4	Kan	2YT	none	3-4 h	37 °C
His_6-hGSTP1-1	pET15-b-His_6-hGSTP1	Amp	2YT	2.5	3-4 h	37 °C
BphK_J2315	pET22-b-BphK_J2315	Amp	2YT	1.0	3-4 h	37 °C
TKB587	pET11-d-TKB587	Amp	2YT	0.25	5-6 h	37 °C
TKB611	pET11-d-TKB611	Amp	2YT	0.25	5-6 h	37 °C

Cells were harvested by refrigerated centrifugation at 5 000 x *g* for 10 min immediately after induction. Cell pellets were stored at -20°C until required.

5.3.2.2. Purification of SjGST, SjGST_Y7F, mGSTA4-4 and BphK_J2315

The expressing pellet was resuspended (4 ml/g wet cell pellet) in buffer E containing one protease inhibitor cocktail tablet (EDTA-free). The resuspended pellet was sonicated for 15 min (30 s on / 30 s off) on ice and was then centrifuged at 27 000 x *g* for 30 min at 4°C to remove insoluble debris. The supernatant was filtered through a 0.45 µm membrane prior to chromatography at 4°C. The cell lysate was loaded onto a 20 ml GSTPrep FF 16/10 column (GE Healthcare) previously equilibrated with binding buffer E. The column was then washed with 5 column volumes of buffer A before elution using a linear gradient of glutathione (0 – 100% buffer F) over 5 column volumes. Fractions were analysed by SDS-PAGE and proteins that did not bind to the column were shown to be in the flow-through fraction with the rest of the unbound material. Fractions containing the protein of interest were pooled, dialysed twice against 4 L of buffer D, using a Spectra/Por® 8,000 molecular-weight-cutoff dialysis tubing, and fractionated into aliquots of 1 mL. 225 µl of sterile 80% glycerol were added to each 1 mL protein aliquot in a 1.5 mL cryotube. Tubes were frozen in liquid nitrogen and stored at -20°C until required.

5.3.2.3. Purification of His_6-SjGST, His_6-hGSTP1-1, His_6-mGSTM1-1, TKB587 and TKB611

The expressing pellets were resuspended in binding buffer A (4 ml per gram of wet cell paste) with one protease inhibitor cocktail tablet (EDTA-free) and disrupted by sonication (15 pulses of 30 s at 30 s intervals) at 4°C. The cell debris were removed by centrifugation at 27 000 x *g* for 30 minutes at 4°C, after which the supernatant was filtered through a 0.45 µm membrane prior to chromatography. The cell lysate was loaded onto a 5 ml HisTrap™ HP column (GE Healthcare) previously

equilibrated with buffer A. The column was then washed with 5 column volumes of buffer A before the bound material was eluted using a linear gradient of imidazole (0 – 100% buffer B) over 20 column volumes at 4°C. Fractions were analysed by SDS-PAGE and those containing the protein of interest were pooled and applied onto a 320 ml HiPrep Sephacryl S-200 HR column (GE Healthcare) pre-equilibrated with buffer C, except for TKB587 which was applied to a 120 ml HiLoad Superdex S-75 column (GE Healthcare). The column was then washed with one column volume of buffer C. Fractions containing the protein of interest were pooled, dialysed twice against 4 L of buffer D, using a Spectra/Por® 8,000 molecular-weight-cutoff dialysis tubing, and fractionated into aliquots of 1 mL. 225 µl of sterile 80% glycerol were added to each 1 mL protein aliquot in a 1.5 mL cryotube. Tubes were frozen in liquid nitrogen and stored at -20°C until required.

5.3.2.4. Sephacryl 200 and Superdex 75 columns calibration

Calibrations were performed with low and high molecular weight calibration kits (GE Healthcare) according to the manufacturer's guide lines. Columns were preequilibrated with 20 mM Tris buffer containing 150 mM NaCl, pH 7.5 and then run with Blue Dextran 2000 (1 mg/ml). Elution volumes were referred as the void volumes, V_0. Mixture of low and high molecular weight standards (5 mg/ml each) were successively run and the elution volumes noted Ve. The gel phase distribution coefficient (Kav) of the protein was calculated as $(Ve - V_0)/(Vc - V_0)$, where Vc is the total volume of the column (320 ml). Kav was plotted against log(Mw) of each sample injected (Figure 5.10. and 5.11.) and a linear equation was fitted by Sigmaplot software (Systat Software Inc.).

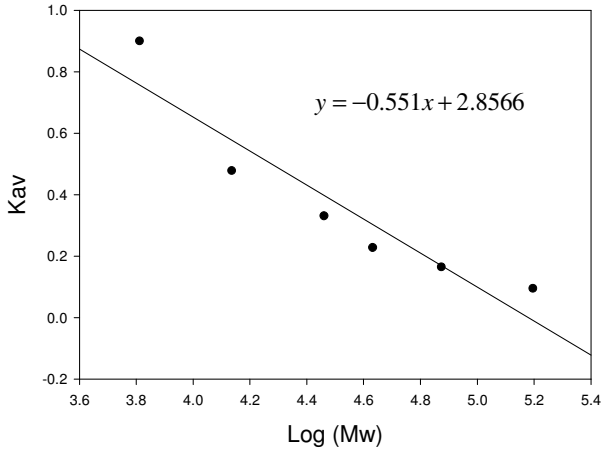


Figure 5.10. Calibration curve of Sephacryl 200 Hiprep 26/60 column.

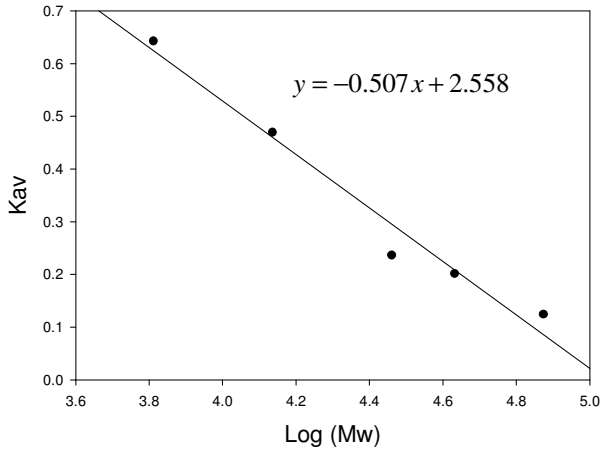


Figure 5.11. Calibration curve of Superdex 75 Hiprep 16/60 column.

5.3.3. Analyses of purified proteins

5.3.3.1. Bradford assay

Protein concentrations of cell lysates, soluble fractions and purified GSTs were determined by the method of Bradford [4] using a Quick Start™ Bradford Protein Assay (Bio-Rad). This method is based on the change in absorbance of Coomassie Blue G-250 upon binding of protein. The dye exists in three forms: cationic (red), neutral (green), and anionic (blue). Under acidic conditions, the dye is

predominantly in the doubly protonated red cationic form (λ_{max} = 470 nm). However, when the dye binds to protein, it is converted to a stable unprotonated blue form (λmax = 595 nm). It is this blue protein-dye form that is detected at 595 nm in the protein assay. A standard curve was produced from serial dilutions (Table 5.1) of a known stock concentration of BSA (2 mg/mL).

1 mL of 1X Dye Reagent was added to a clean, dry cuvette containing the protein sample (20 μL). The solution was mixed by inverting the cuvette 3-4 times and incubated at room temperature for 10 minutes. Each sample was carried out in duplicate. The absorbance of the standards and unknown samples was measured at 595 nm. The concentration of protein was calculated according to the standard curve.

Standard	**Concentration (mg/mL)**
A	2.00
B	1.50
C	1.00
D	0.75
E	0.50
F	0.25
G	0.125
H	0.025
I	0.00

Table 5.1. Standard curve concentrations.

5.3.3.2. Liquid chromatography-mass spectrometry (LC-ESI-MS)

LC-ESI-MS was performed on a MicroMass Platform II quadrupole mass spectrometer equipped with an electrospray ion source. The spectrometer cone voltage was ramped from 40 to 70 V and the source temperature set to 140°C. Protein samples were separated on a Jupiter C5 reverse phase column (5 μm, 250 x 4.6 mm,

Phenomenex) with a Waters HPLC 2690 directly connected to the spectrometer. Proteins were eluted from the column with a 5-95% acetonitrile (containing 0.01% TFA) gradient at a flow rate of 0.1 ml.min^{-1}. The total ion count in the range 400-1980 m/z was scanned at 0.1 s intervals. The scans were accumulated, spectra combined and the molecular mass determined by the MaxEnt and Transform algorithm of the Mass Lynx software (Micromass, U.K.).

5.3.3.3. MALDI-TOF Mass Spectrometry

Proteins were buffer-exchanged against a Tris-HCl buffer (20 mM) using a 10000 MWCO centrifuge filter (Vivaspin) and co-crystallised with sinapinic acid used as a matrix, directly on MALDI plate. A volume of 0.5 µl was deposited on a MALDI plate and was kept at room temperature until dried. MALDI-MS spectra (two replicates) were measured on a Voyager DE-STR MALDI-TOF mass spectrometer. Spectra were acquired in the range of 20,000 to 65,000 Da. Other variables were set as follows: accelerating voltage = 25,000 V, grid voltage = 96%, guide wire = 0.15%, extraction delay time = 800 nsec. Typically, 100 shots were taken for each sample combined and analysed using the DataExplorer software from Applied Biosystems.

5.3.3.4. *In vitro* glutathione transferase assays

Activity assays were performed on a Varian's Cary 50 UV-Vis spectrophotometer and inhibition assays were carried out on a Molecular Devices SpectraMax M5 UV-visible 96 well-plates spectrophotometer.

5.3.3.4.1. Activity assays

The initial rates of GST-catalyzed conjugation of GSH with CDNB were determined spectrophotometrically according to the method of Habig *et al.* [5]. The K_m^{CDNB} for the CDNB-GSH conjugation reaction was determined using a range of

CDNB concentrations while the GSH concentration was fixed at 2 mM. Determination of the K_m^{GSH} was conducted using a range of GSH concentrations, while the CDNB concentration was constant (1 mM final concentration). Enzymatic reactions were carried out in potassium phosphate, 0.1 M, pH 6.8 (reaction buffer) containing 2.5% ethanol.

Preparation of stock solutions

GST samples and a GSH stock solution (40 mM) were prepared in reaction buffer. A CDNB stock solution (20 mM) was prepared in 50% ethanol.

A set of 12 substrate concentrations was prepared as a two-fold serial dilution set for each substrate as follows.

- GSH:

1 mL of GSH stock solution (40 mM) was added to 9 mL of reaction buffer, giving a second GSH stock solution of 4 mM. 5.6 mL of this second stock solution were added to 4.4 mL of reaction buffer, giving a GSH concentration of 2222.2 µM.
From this solution, 5 mL were serially added 11 times to 5 mL of reaction buffer, giving the concentrations presented in the table below.

- CDNB:

1 mL of CDNB stock solution (20 mM) was added to 9 mL reaction buffer, giving a second CDNB stock solution of 2 mM.
5.6 mL of this second stock solution were added to 4.4 mL of reaction, giving a CDNB concentration of 1111.1 µM in reaction buffer containing 2.8% ethanol.
From this solution, 5 mL were serially added 11 times to 5 mL of reaction buffer with 2.8% ethanol, giving the concentrations presented in the table below.

GSH initial concentrations (μM)	GSH concentrations in the reaction (μM)	CDNB initial concentrations (μM)	CDNB concentrations in the reaction (μM)
2222.2	2000	1111.1	1000
1111.1	1000	555.6	500
555.6	500	277.8	250
277.8	250	138.3	125
138.3	125	69.4	62.5
69.4	62.5	34.7	31.3
34.7	31.3	17.4	15.6
17.4	15.6	8.7	7.8
8.7	7.8	4.4	3.9
4.4	3.9	2.2	2
2.2	2	1.1	1
1.1	1	0.51	0.5

Lower substrate concentrations were used when assaying SjGST_Y7F:

GSH initial concentrations (μM)	GSH concentrations in the reaction (μM)	CDNB initial concentrations (μM)	CDNB concentrations in the reaction (μM)
22.2	20	88.8	80
11.1	10	44.4	40
5.6	5	22.2	20
2.8	2.5	11.1	10
1.4	1.25	5.6	5
0.7	0.625	2.8	2.5
0.35	0.31	1.4	1.25
0.17	0.15	0.7	0.625
0.085	0.08	0.35	0.31
0.042	0.04	0.17	0.15

Reaction settings

The reactions were maintained at 25°C, and the conditions were generally saturating for the invariant substrate.

To a 1 ml cuvette were added:

- 900 µl of either CDNB (8.7 to 1111.1 µM) or GSH (17.4 to 2222.2 µM);
- GST (50 µl, ~0.15 $mg.ml^{-1}$);

The solution was mixed well and after incubation at 25°C for 5 minutes, GSH or CDNB (50 µl, 40 or 20 mM, respectively) was added quickly and mixed. The enzymatic reactions were monitored at 340 nm ($\Delta\varepsilon = 9.6\ mM^{-1}.cm^{-1}$) for 5 minutes.

Each reaction was carried out in triplicate. Controls were performed without enzyme for each substrate concentration. The specific activity of the enzymes was expressed as micromoles of substrate per minute per milligram of enzyme and was corrected for the rate of the spontaneous non-enzymatic conjugation reaction of CDNB and GSH. Kinetic data were analysed using the graphing software SigmaPlot®.

5.3.3.4.2. Inhibition studies

Inhibition studies were carried out on SjGST and hGST P1-1. In these experiments, enzymatic reactions were carried out in potassium phosphate, 0.1 M, pH 6.8 (reaction buffer) containing 10% DMSO.

GST samples and a stock solution of GSH (40 mM) were prepared in reaction buffer. Stock solutions CDNB (20 mM) and inhibitors (20 mM) were prepared in DMSO.

Each reaction was carried out in triplicate. Controls were performed in the same conditions without enzyme and initial rates were corrected for the rate of the spontaneous non-enzymatic conjugation of CDNB and GSH. Data were collected and analysed using SigmaPlot® software.

Measure of IC_{50}s

IC_{50} values, with respect to the GSH-CDNB conjugation reaction, were determined for SjGST assayed with compounds **4**, **5a** to **5j**; and hGST P1-1 assayed with compounds **4**, **5c**, **5g**, **5h** and **5i**.

Sets of 17 inhibitor concentrations were prepared in DMSO, spanning a range from 0.2 to 1,000 μM, final concentrations:

Inhibitor initial concentrations (μM)	4	8	12	16	20	40	80	120	160	200	400
Inhibitor final concentrations in the reaction (μM)	0.2	0.4	0.6	0.8	1.0	2	4	6	8	10	20

Inhibitor initial concentrations (μM)	800	1200	1600	2000	3000	4000	10,000	20,000
Inhibitor final concentrations in the reaction (μM)	40	60	80	100	150	200	500	1,000

To a 360 μl well were added 240 μL of reaction buffer, GST (15 μL, ~ 0.15 $mg.mL^{-1}$) and inhibitor (15 μL, 17 concentrations spanning from 4 μM to 20 mM). The solution was mixed well and incubated at 25°C for 5 minutes. CDNB (15 μL, 20 mM) and GSH (15 μL, 40 mM) were added and the plate was quickly shaked. Absorbance was measured at 340 nm, 25°C for 5 minutes.

Measure of K_is

The K_i values with respect to both GSH and CDNB were determined for SjGST and hGST P1-1 with compounds **5c** and **5g**.

Solutions of CDNB were prepared in 50% DMSO.

Sets of 8 substrate concentrations were serially prepared from the stock solutions, giving the following concentrations:

GSH initial concentrations (mM)	GSH concentrations in the reaction (μM)	CDNB initial concentrations (mM)	CDNB concentrations in the reaction (μM)
40	2000	20	1000
20	1000	10	500
10	500	5	250
5	250	2.5	125
2.5	125	1.25	62.5
1.25	62.5	0.625	31.3
0.625	31.3	0.313	15.6
0.313	15.7	0.157	7.8

Sets of 5 inhibitor concentrations were used spanning a range from 2.5 to 200 μM (final concentrations) as a three-fold serial dilution set. Dilutions were made directly in the 96-well plate. 340 μL of a 235 μM stock solution prepared in reaction buffer with 6% DMSO were pipetted into 4 wells (3 replicates + blank). Subsequently, 85 μL were serially added 4 times to 170 μL of reacion buffer containing 6% DMSO, giving the following concentrations:

Initial inhibitor concentrations (μM)	235	78.4	26.1	8.7	2.9
Final inhibitor concentrations in the reaction (μM)	200	66.7	22.2	7.4	2.5

To the wells containing inhibitor (255 μL) were added 15 μL of GST (~0.15 $mg.mL^{-1}$) and the solution was incubated at 25˚C for 5 minutes. Subseqently, 15 μL of

GSH (0.313 to 4 mM) and 15 µl of CDNB (0.157 to 2 mM) were added to the reaction. The plates were quickly shaked and absorbance was measured at 340 nm, 25˚C for 5 minutes.

5.3.3.4.3. Isothermal Calorimetry

Isothermal calorimetry (ITC) was carried out in collaboration with Prof. Alan Cooper and the Glasgow Biological Microcalorimetry Facility. The ITC measurements were performed on a VP-ITC calorimeter (Microcal Inc Northampton, USA) at 25°C.

SjGST and hGST P1-1 were dialysed against a 0.1 M potassium phosphate buffer, pH 6.8 containing 10 % DMSO. The concentrations of SjGST and hGSTP1-1 (~10 µM for titrations with **5c** and **5g**; ~20 µM for titrations with **4**) were determined by measuring absorption at 280 nm [$\Delta\varepsilon_{280}$(SjGST) = 41.2 $mM^{-1}.cm^{-1}$; $\Delta\varepsilon_{280}$(hGST P1-1) = 38.8 $mM^{-1}.cm^{-1}$].

Compounds **5c**, **5g** and **4** were provided as solids by Venughopal Bhat (University of Edinburgh). The right amounts of compounds were accurately weighed out using an AnD FX-200 benchtop balance and dissolved in the same dialysis buffer. Ligand concentrations were approximately 600 µM.The same buffer was used for titration, instrument calibration and baseline controls.

The proteins SjGST and hGST P1-1 were placed in the 2 ml sample chamber and **5c**, **5g** or **4** in the syringe. A typical ITC measurement consisted of a first control injection of 1 µl followed by 29 successive injections of 10 µl for 20 s with a 3 minutes interval between each injection.

Control experiments in which ligands were directly injected in the buffer without enzyme were performed in order to evaluate the heat contributions due to coupled protonation events upon binding. The observed heat effects were identical to the heat signals after complete saturation of the proteins. Therefore, the non-specific background was usually estimated by averaging the small heats at the end of the SjGST and hGST P1-1 titrations.

Raw data were collected and the baseline was corrected for ligand heats dilution. The peaks generated were integrated using ORIGIN software (Microcal Inc) by plotting the values in microcalories against the molar ratio of injectant to reactant within the cell.

Data were fitted using the one single-site binding model. From the dissociation constant K_D and the reaction enthalpy value ΔH, the change in free Gibbs energy ($\Delta G°$) and entropy change ($\Delta S°$) can be calculated using the equation $\Delta G° = -RT \ln(1/K_D) = \Delta H - T\Delta S°$ where R is the universal gas constant and T the absolute temperature.

5.4. Molecular docking

5.4.1. Ligand alignments

The superimposition of glutathione S-transferase ligands was carried out by using Relibase+3.0.0 [6]. First, a search was performed to find binding sites that share a sequence identity between 40% and 100% to the target SjGST crystal structure 1M9A [7]. The 38 structures with bound ligand were superimposed by using binding site residues only. Finally, the ligands from the superimposed structures were extracted and analysed.

5.4.2. Binding mode prediction of SjGST with **5g**

The SjGST crystal structure 1M9A [8] was used as initial conformation for binding mode generation. The glutathione group of **5g** was mapped onto the 1M9A ligand coordinates. The thiophen hydrazone group of **5g** was oriented towards the cavity, lying at the end of the S-hexyl site, as observed for the ligand bound to cGSTM1-1 with pdb code 1C72 [7]. In the next step, the protein in complex with **5g** was minimized, considering the ligand as fully flexible. For the protein all residues were kept rigid, except the amino acids that define the pocket at the end of the S-hexyl site (R103, V106, V161, V162, Q204).

5.4.3. Binding mode prediction of hGST P1-1 with **5c**

The hGST P1-1 crystal structure 6GSS [9] was used as initial conformation for binding mode generation. The glutathione group of **5c** was mapped onto the 6GSS ligand coordinates. The tertbutyl hydrazone fragment of **5c** was oriented towards the cavity, lying at the end of the S-hexyl site, as observed for the ligand bound to cGSTM1-1 with PDB code 1C72 [7]. In the next step, the protein in complex with **5c** was minimized, considering the ligand as fully flexible. For the protein all residues were kept rigid, except the amino acids that define the pocket at the end of the S-hexyl site (R100, Y103, I161, H162, N204).

5.5. NMR analyses on **3f** and **3g**

^{1}H NMR spectra were recorded in DMSO on a Brüker dpx600 (600 MHz) instrument and calibrated to residual solvent peaks ($CDCl_3$ 7.26 ppm). The data are presented as follows: chemical shift (in ppm on the δ scale), multiplicity (s=singlet, d=doublet), the coupling constant (J, in Hertz) and integration.

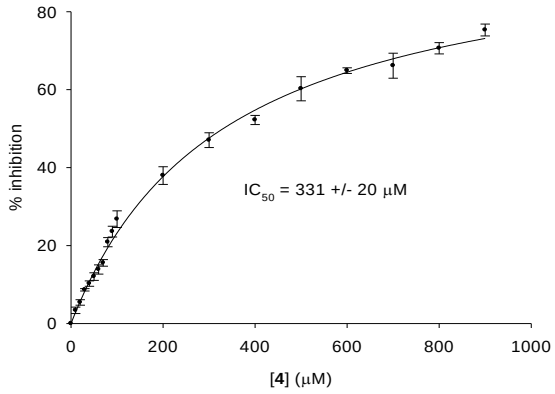
5.6. Chapter 5 references

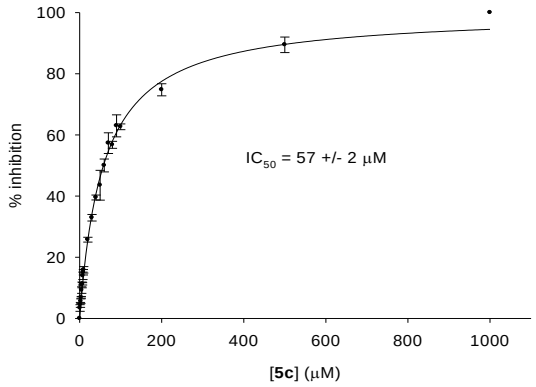
1. Chang, M., J.L. Bolton, and S.Y. Blond, *Expression and purification of hexahistidine-tagged human glutathione S-transferase P1-1 in Escherichia coli.* Protein Expr Purif, 1999. **17**(3): p. 443-8.
2. Jeppesen, M.G., et al., *The crystal structure of the glutathione S-transferase-like domain of elongation factor 1Bgamma from Saccharomyces cerevisiae.* J Biol Chem, 2003. **278**(47): p. 47190-8.
3. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.
4. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal Biochem, 1976. **72**: p. 248-54.
5. Habig, W.H., M.J. Pabst, and W.B. Jakoby, *Glutathione S-transferases. The first enzymatic step in mercapturic acid formation.* J Biol Chem, 1974. **249**(22): p. 7130-9.
6. Bergner, A., et al., *Use of Relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects.* Biopolymers, 2001. **61**(2): p. 99-110.
7. Chern, M.K., et al., *Tyr115, gln165 and trp209 contribute to the 1, 2-epoxy-3-(p-nitrophenoxy)propane-conjugating activity of glutathione S-transferase cGSTM1-1.* J Mol Biol, 2000. **300**(5): p. 1257-69.
8. Cardoso, R.M., et al., *Characterization of the electrophile binding site and substrate binding mode of the 26-kDa glutathione S-transferase from Schistosoma japonicum.* Proteins, 2003. **51**(1): p. 137-46.
9. Oakley, A.J., et al., *The structures of human glutathione transferase P1-1 in complex with glutathione and various inhibitors at high resolution.* J Mol Biol, 1997. **274**(1): p. 84-100.

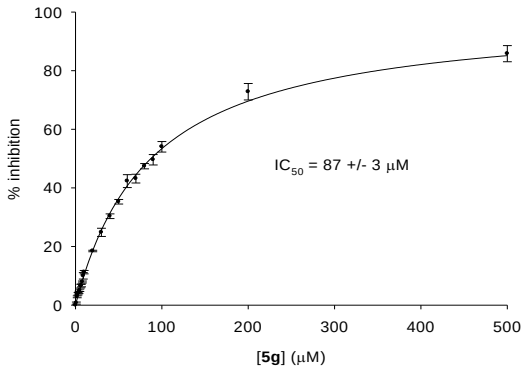
Appendices

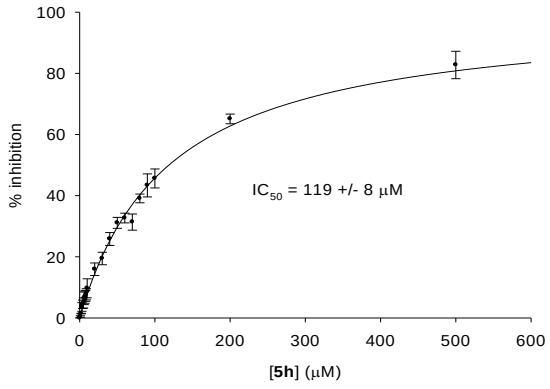
IC_{50} data

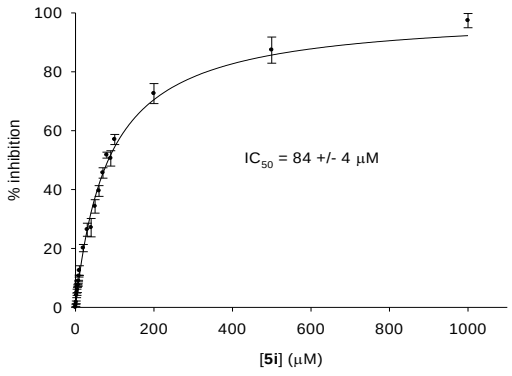
hGST P1-1 assayed with **4**, **5c**, **5g**, **5h** and **5i**:



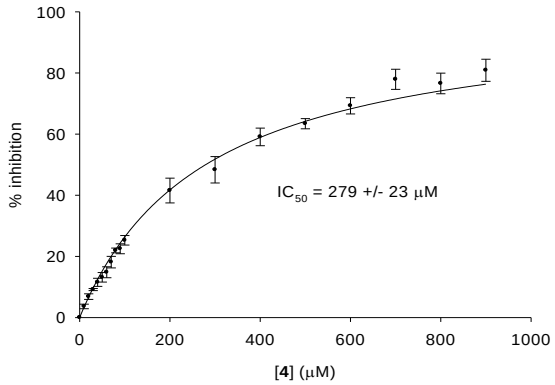


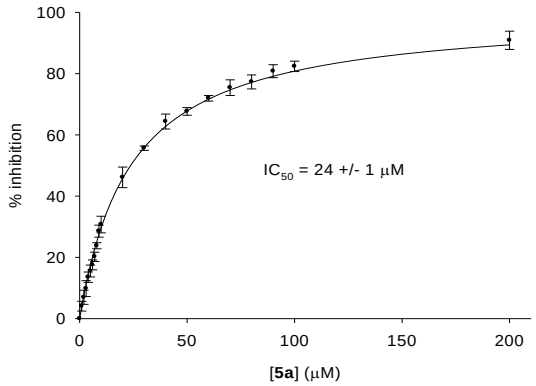


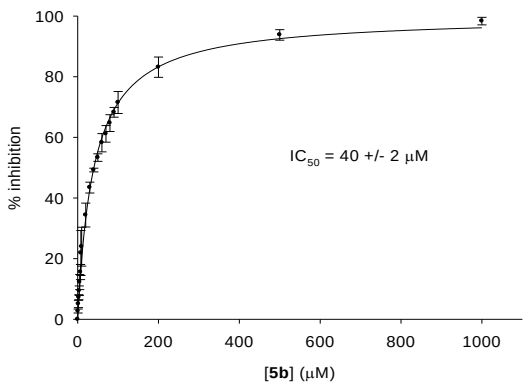


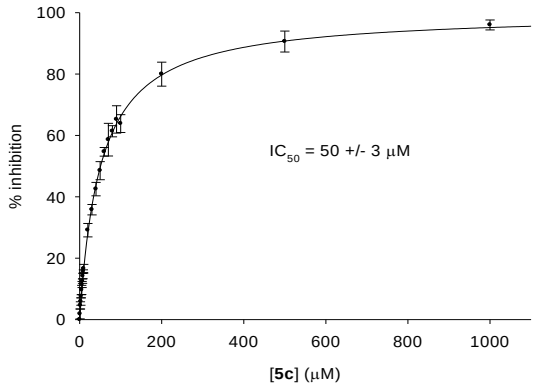


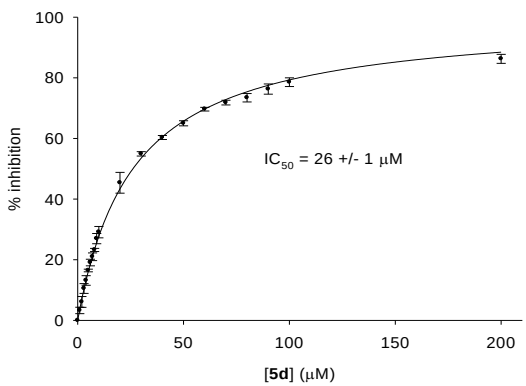
SjGST assayed with **4** and **5a** to **5j**:

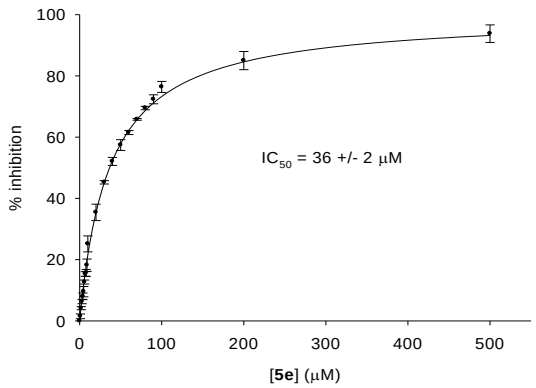


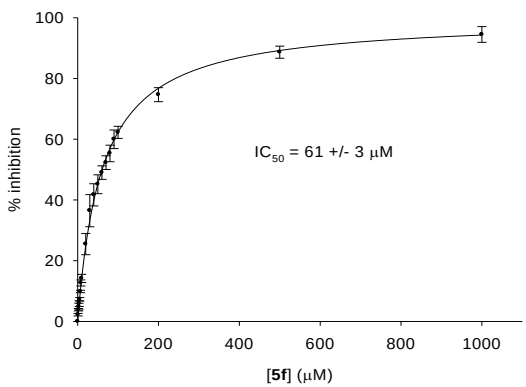


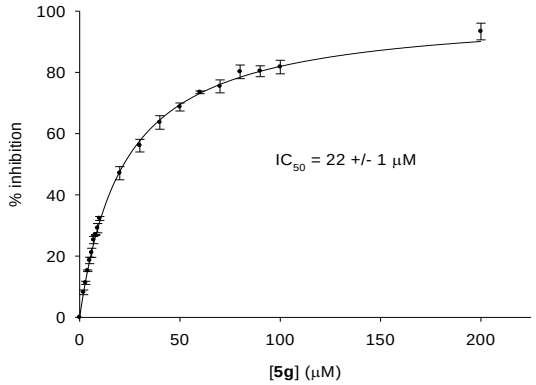


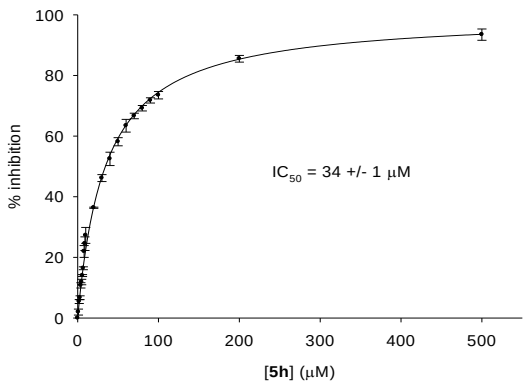


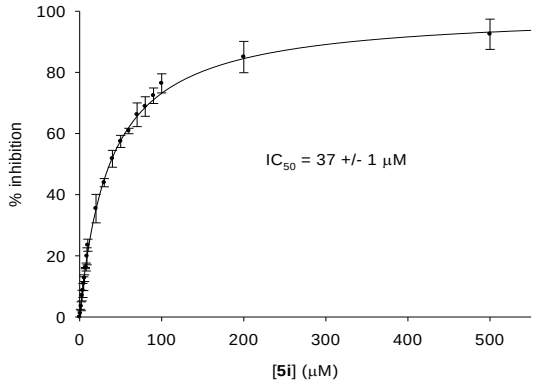


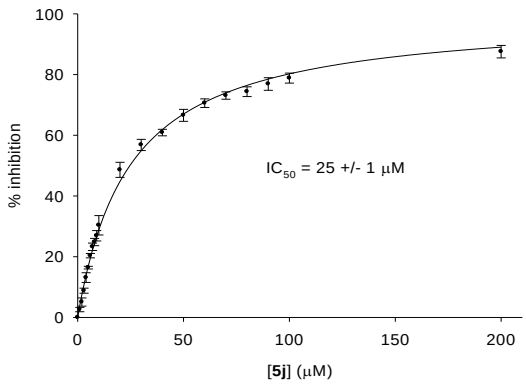





% inhibition
IC50 = 22 +/- 1 μM
[5g] (μM)

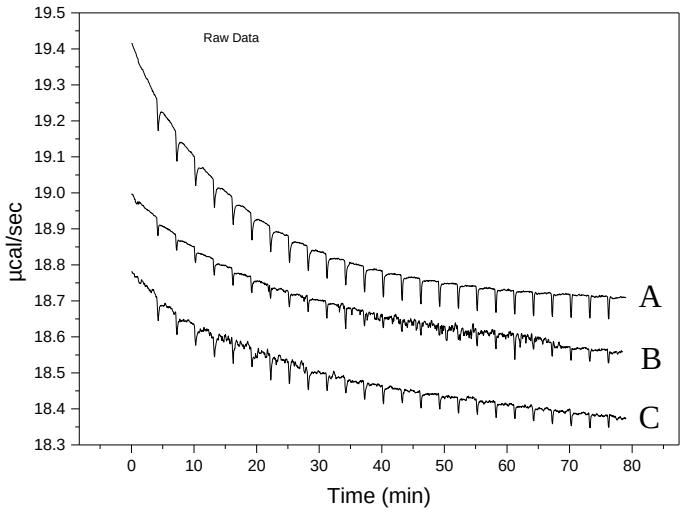

% inhibition
IC50 = 34 +/- 1 μM
[5h] (μM)


% inhibition
IC50 = 37 +/- 1 μM
[5i] (μM)

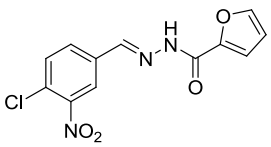

% inhibition
IC50 = 25 +/- 1 μM
[5j] (μM)

Raw data of ITC controls

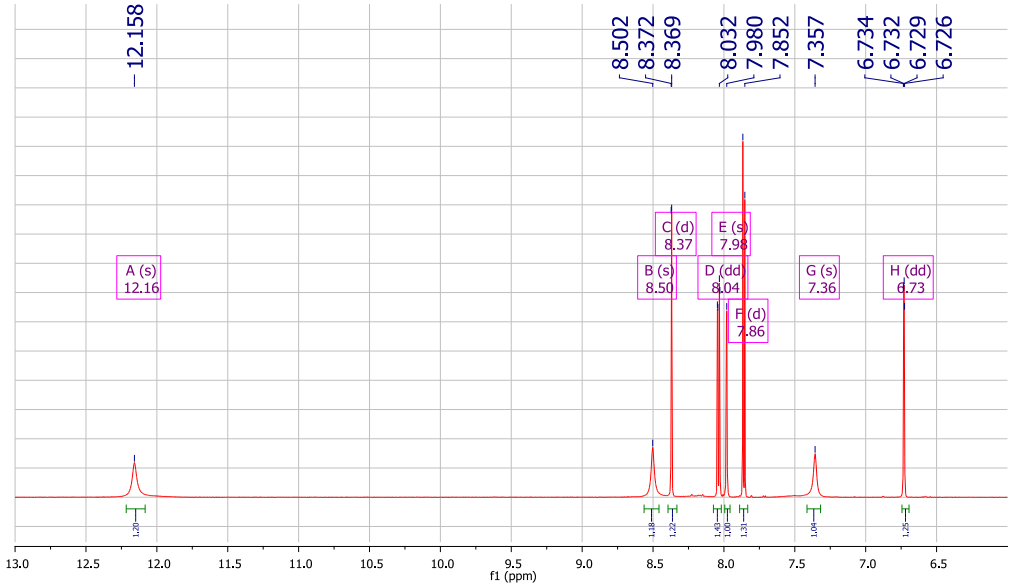
ITC control experiments: heat signals generated by injection of a ligand (~600 μM; A: **5c**; B: **5g**; C: **4**) directly into buffer without enzyme. The observed heat effects were identical to the heat signals after complete saturation of the proteins

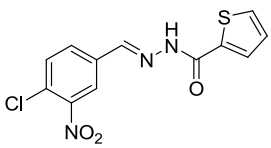


^{1}H NMR assignments

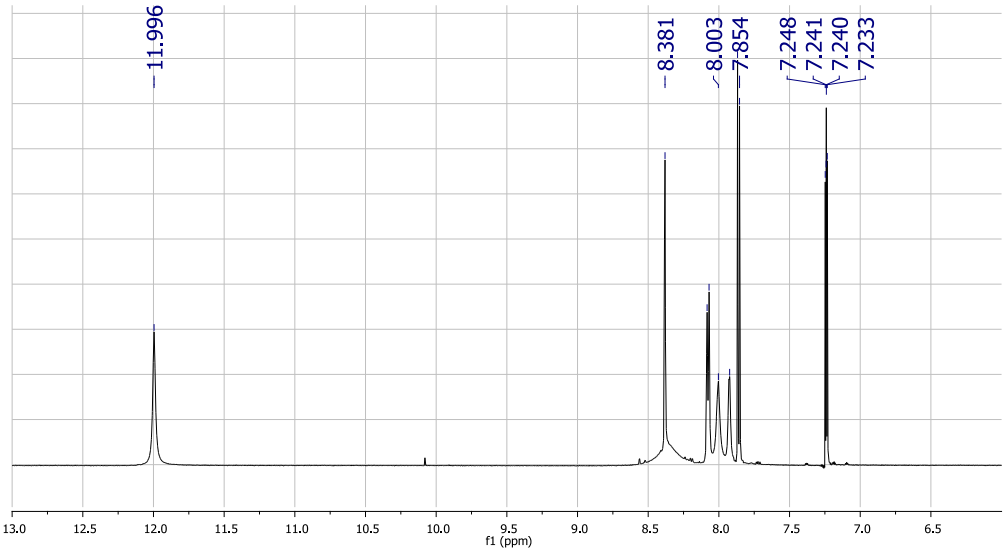


Hydrazone **3f**: ^{1}H NMR δ_H 12.16 (s, 1H), 8.50 (s, 1H), 8.37 (d, *J* = 1.96 Hz, 1H), 8. 37 (dd, *J* = 8.45, 2.02 Hz, 1H), 8.03 (d, *J* = 0.98 Hz, 1H), 7.85 (d, *J* = 8.42 Hz, 1H), 7.36 (s, 1H), 6.73 (dd, *J* = 3.52, 1.74 Hz, 1H).





Hydrazone **3g**: ^{1}H NMR δ_H 12.00 (s, 1H), 8.38 (s, 1H), 8.00 (d, *J* = 1.96 Hz, 1H), 7.08 (dd, *J* = 8.45, 2.02 Hz, 1H), 8.07 (d, *J* = 0.98 Hz, 1H), 7.85 (d, *J* = 8.42 Hz, 1H), 7.24 (s, 1H), 7.24 (dd, *J* = 3.52, 1.74 Hz, 1H).



nature chemistry

ARTICLES

PUBLISHED ONLINE: 16 MAY 2010 | DOI: 10.1038/NCHEM.658

Nucleophilic catalysis of acylhydrazone equilibration for protein-directed dynamic covalent chemistry

Venugopal T. Bhat[1†], Anne M. Caniard[1†], Torsten Luksch[2], Ruth Brenk[2], Dominic J. Campopiano[1*] and Michael F. Greaney[1*]

Dynamic covalent chemistry uses reversible chemical reactions to set up an equilibrating network of molecules at thermodynamic equilibrium, which can adjust its composition in response to any agent capable of altering the free energy of the system. When the target is a biological macromolecule, such as a protein, the process corresponds to the protein directing the synthesis of its own best ligand. Here, we demonstrate that reversible acylhydrazone formation is an effective chemistry for biological dynamic combinatorial library formation. In the presence of aniline as a nucleophilic catalyst, dynamic combinatorial libraries equilibrate rapidly at pH 6.2, are fully reversible, and may be switched on or off by means of a change in pH. We have interfaced these hydrazone dynamic combinatorial libraries with two isozymes from the glutathione *S*-transferase class of enzyme, and observed divergent amplification effects, where each protein selects the best-fitting hydrazone for the hydrophobic region of its active site.

Dynamic covalent chemistry (DCC) uses reversible chemical reactions to set up equilibrating assemblies of molecules at thermodynamic equilibrium[1–4]. The resultant dynamic combinatorial library (DCL) is responsive to the addition of a template, which will selectively amplify the best binding compounds from the equilibrium distribution. The essence of the concept lies in the subsequent adjustment of the DCL equilibrium, which will express more of the best binding compounds at the expense of the poorer ones. A DCL is thus adaptive and capable of evolutionary behaviour, whereby individual components are either amplified or reduced in response to template-directed binding events. These concepts have been applied to diverse problems in biological and medicinal chemistry[5–11], synthetic receptor–ligand interactions[12–16], self-replication[17–19], complex molecule synthesis[20–22] and materials science[23,24]. Taken together, they represent the best characterized examples to date of systems chemistry, which looks to synthesize complex molecular networks and study their properties and behaviour in macrocosm, rather than as a sum of their individual components[25,26].

We are interested in DCC systems that use a biological molecule, such as a protein, to template assemblies of small molecules at dynamic equilibrium[27]. Here, the DCC experiment provides a method for discovering, studying and ranking novel protein ligands, concepts fundamental to medicinal chemistry. In these terms, the DCC process bridges the gap between targeted chemical synthesis of drug candidates and their biological binding assay, meshing the two processes into a single step in which the structure of the biological target directs the assembly of its own best inhibitor *in situ*.

A particular challenge for DCC in biological systems lies in the implementation of a suitable reversible reaction that can operate effectively under the physiological conditions required by the biotemplate. Lehn has defined two limiting cases for DCL construction: adaptive and pre-equilibrated DCC[28]. The adaptive DCL represents the ideal scenario, where the DCL chemistry is fully compatible with the biological target and the ensuing binding events control the evolution of the DCL composition. Pre-equilibrated DCL refers to the cases where the reversible chemistry used to constitute the

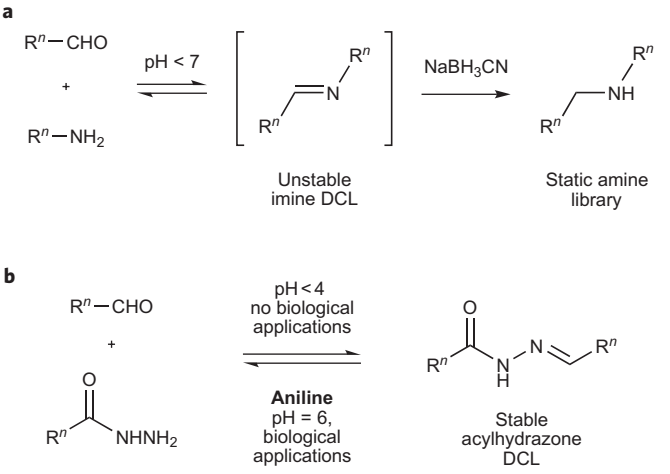


Figure 1 | Transimination reactions for DCC. a, Imine DCLs: reversible addition of amines to aldehydes gives unstable imines that cannot be isolated or analysed directly, necessitating an *in situ* reduction step. The resultant static library of amines may or may not share the binding profile of the imine precursors. **b**, Acyl hydrazone DCLs: reaction of aldehydes with hydrazides gives acylhydrazones that have good stability and are amenable to analysis. Equilibration requires acidic conditions that are incompatible with biological targets—a nucleophilic catalyst such as aniline may enable DCL formation at biocompatible pH.

[1]EastChem, School of Chemistry, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JJ, UK, [2]College of Life Sciences, University of Dundee, James Black Centre, Dow Street, Dundee DD1 5EH, UK; [†]These authors contributed equally to this work. *e-mail: Dominic.Campopiano@ed.ac.uk; Michael.Greaney@ed.ac.uk

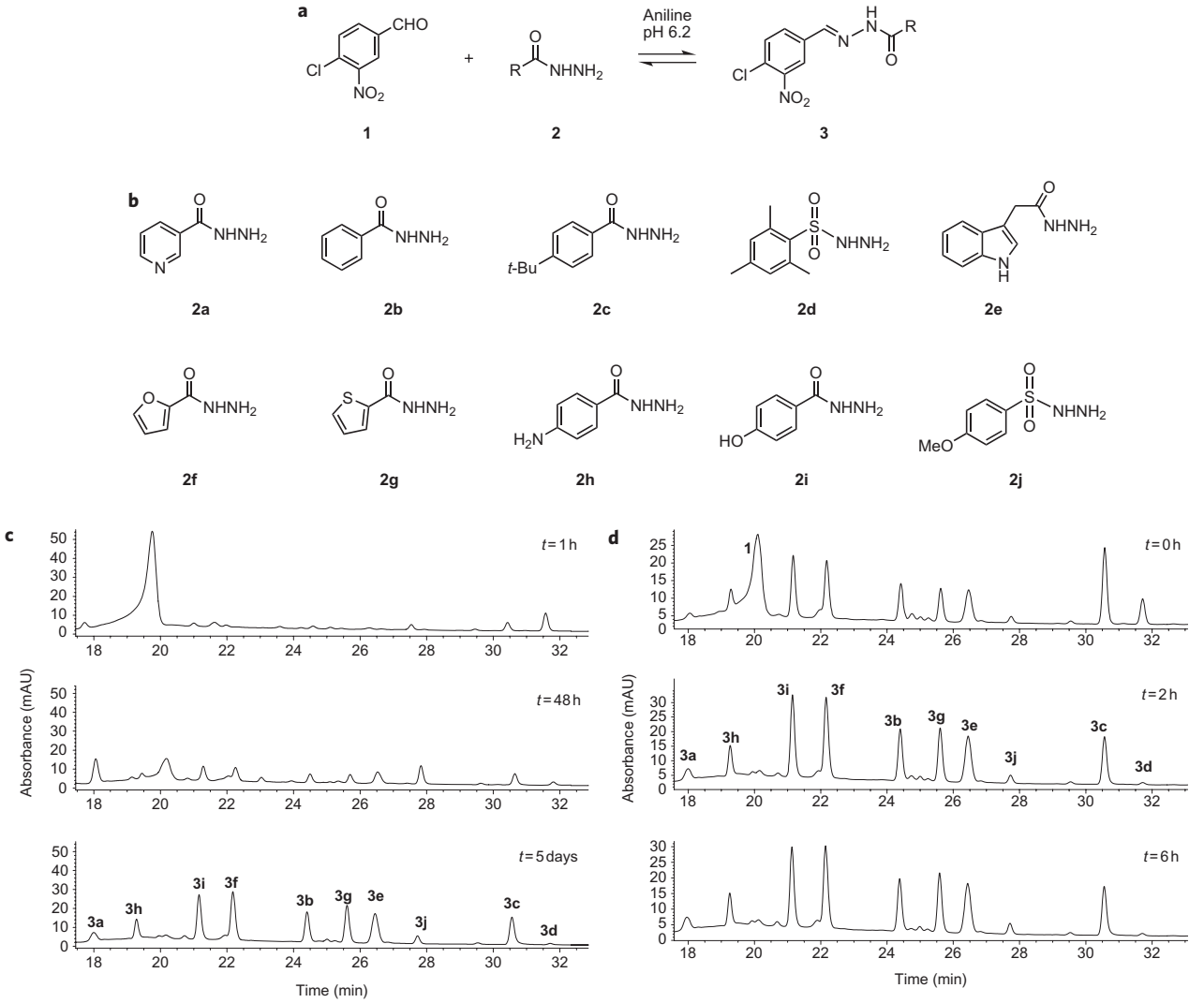


Figure 2 | Aniline-catalysed acylhydrazone formation. a, Aldehyde equilibration with hydrazide to form an acylhydrazone. **b**, Hydrazide components of the ten-membered DCL. **c**, DCL established in the absence of aniline. Conditions: aldehyde (5 μM), hydrazides (20 μM each) in NH_4OAc buffer (50 mM, pH = 6.2) containing 15% DMSO. **d**, DCL established in the presence of aniline (10 mM).

DCL is not compatible with the biological target, meaning that the DCL and the target must be separated in some manner. This results in static libraries in which the molecular recognition events that control DCL composition are lost. Given the challenges associated with conducting fast, freely reversible chemistry under physiological conditions, it is not surprising that methods for true adaptive DCL generation are limited, with the majority of successful systems using sulfur-based transformations such as disulfide bond formation or thiol conjugate addition[29–32]. The development of new methods for adaptive DCLs is thus central to the application of DCC to biological systems, as the chemistry will define the target scope and range of available DCL components.

The reversible formation of C=N imine-type linkages emerged early on as a DCL-forming reaction[33]. The ready availability of diverse carbonyl and amine building blocks, plus the extensive precedent of imine formation in biochemical systems, makes it an ideal candidate reaction. However, the inherent instability of imines in aqueous solution presents serious analytical and isolation problems in the DCC context. The solution to this in the field of biological DCC has been to construct pseudo-adaptive DCLs where the imine linkage is reduced *in situ* to an amine with an external hydride source. The resulting library contains static amine components that can correspond to the imines in binding affinity, although both false-positive and false-negative results are possible. In addition, the introduction of an *in situ* reduction step complicates the DCL equilibration and makes it difficult to distinguish between genuine thermodynamic selection of the best binders and selection of those compounds that are kinetically favoured.

An advance on simple imine formation in DCC came from the Sanders group, who introduced acylhydrazones as reversible linkages[34]. The reaction has proven to be an excellent balance between facile reversibility and product stability; the acylhydrazone products formed are stable to analysis and isolation, and the reaction has very good equilibration properties, as is made evident by its application to a large number of elegant abiological DCC studies subsequently reported by the Sanders group[35–37]. It has not, however, been generally possible to apply this reaction directly to adaptive biological DCC systems because of the acidic pH required for reversibility to occur in a reasonable timeframe ($pH < 4$)[38,39]. A single elegant study from Poulsen has shown that slow equilibration of acylhydrazones, taking one week at pH 7.2, can be accelerated in the presence of the enzyme carbonic anhydrase, enabling *in situ* identification of binders using mass spectrometry[40]. We were keen to apply this proven reaction to our DCC studies of enzymes, and reasoned that it could be harnessed as a powerful tool for biological investigation if a suitable catalyst

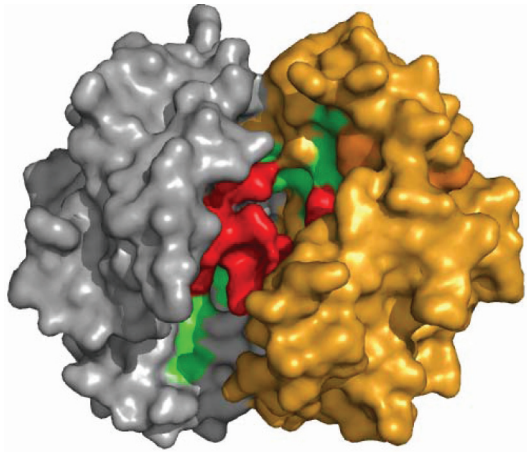


Figure 3 | Structure of GST illustrating H- and G-sites. Grey, monomer 1; yellow, monomer 2; green, G-site; red, H-site.

could be found to accelerate the equilibration. Nucleophilic catalysis of semicarbazone formation using aniline derivatives was established in classic work from Jencks in the 1960s, and recently applied to hydrazone and oxime formation in peptide ligation systems by Dawson[41–44]. We reasoned that an additive such as aniline could promote the equilibration of acylhydrazides and aldehydes at pH values closer to the physiological window required by biological targets in DCC (Fig. 1).

We began by reacting aldehyde **1** (Fig. 2), related to the known glutathione *S*-transferase (GST) substrate chlorodinitrobenzene (CDNB, see below), with an excess of the ten aryl hydrazides **2a**–**2j** at room temperature. The hydrazides were chosen to randomly display aryl and heteroaryl groups and featured eight acyl and two sulfonyl hydrazides (**2d** and **j**). Equilibration at pH 6.2 was slow, and only two of the ten possible hydrazones could be observed by high-performance liquid chromatography (HPLC) after 1 h (Fig. 2c). Notably, there was a significant amount of free aldehyde **1** present throughout the reaction, despite the presence of excess amounts of the ten different hydrazides. Equilibrium was not complete after 48 h, and required incubation for a further 5 days until the library composition reached a steady-state composition with signals for each of the ten hydrazones **3a**–**3j** being clearly identified. In contrast, repeating the experiment in the presence of excess aniline produced a far higher rate of equilibration. A distribution of acylhydrazones was observed after initial mixing and HPLC sampling, and complete equilibration of the ten components was observed after just 6 h (Fig. 2d). Aldehyde **1** could not be detected following initial mixing, indicating that it was continually being sequestered as an acylhydrazone component, reflecting the faster exchange processes operating in the presence of aniline (see Supplementary Information for a study on the effect of varying aniline concentration on rates of hydrazone formation). We demonstrated the reversibility of the DCL by generating it from a different starting composition, hydrazone **3g** plus the nine other hydrazides and aniline. An identical equilibrium distribution to Fig. 2 was observed, indicating true thermodynamic equilibrium. A second control experiment confirmed the reversibility of the DCL through the addition of excess hydrazide **2b** to the pre-equilibrated DCL, which resulted in a large amplification of the corresponding acylhydrazone **3b** (see Supplementary Information).

Having established that aniline could act as an effective nucleophilic catalyst for hydrazone DCC formation at both a pH and timeframe reasonable for biomolecule stability, our next step was to introduce proteins to the DCL. Our target chosen for DCC interrogation was the GST enzyme superfamily[45]. The GSTs are responsible for cell detoxification, catalysing the conjugation of glutathione (GSH) to a wide variety of xenobiotic electrophiles, thereby protecting the cell from cytotoxic and oxidative stress. We have previously developed thiol conjugate addition DCLs directed towards GST inhibition, and successfully interfaced the enzyme with small molecules so that it controlled library evolution[27]. The GSTs are well suited to exploration using DCC methods, being well-characterized, robust proteins having nascent medicinal chemistry application[46,47]. There are relatively few ligands reported in the literature for GST binding—a plus point, as it would enable us to use DCC as a genuine discovery tool for new binding motifs, rather than as a proof-of-principle process for confirming the binding ability of known ligands. The cytoplasmic GSTs are inherent dimers with active sites composed of residues from both monomers, bifurcating between a highly conserved G-site, which binds the endogenous ligand GSH, and an H-site, which binds hydrophobic substrates for GSH conjugation (Fig. 3). This bisubstrate architecture is particularly appropriate for DCC interrogation, given that the method essentially uses a reversible linkage to couple two sets of fragment structures together[48]. Furthermore, within the GST superfamily, the large, heterogeneous H-sites are functionally evolved to accommodate many different hydrophobic substrates for conjugation, a classically difficult architecture to investigate using orthodox structure-based drug-design methods.

We prepared two recombinant GST isozymes as targets, SjGST from the helminth worm *Schistosoma japonicum*, a drug target in tropical disease[49], and hGST P1-1, a human isoform that has been targeted in the treatment of chemotherapy drug resistance[50]. An initial control experiment with SjGST established that the enzyme retained GSH conjugation activity in the presence of aniline (up to 20 mM). The acylhydrazone DCL prepared in Fig. 2 was then interfaced with the two protein targets and amplification was measured (Fig. 4). Both DCLs demonstrated strikingly clear amplification of hydrazone components; thiophene acylhydrazone **3g** was selected by SjGST and *t*-butylphenyl hydrazone **3c** by hGST P1-1.

Synthesizing the DCL in the presence of bovine serum albumin (BSA, 1 equiv.) as a control experiment produced no measurable amplification of any component, indicating the GST enzymes as being responsible for component amplification. We further demonstrated that amplified components were bound in the target H-region of the active site of the enzyme by performing conjugation experiments with GSH. Conjugation of GSH to the aryl chloride group in hydrazones **3** by means of S_NAr substitution is a slow reaction at pH 6.2, taking several days. In the presence of catalytic amounts of SjGST, however, rapid formation of the S_NAr conjugation adduct for hydrazone **3g** was observed at pH 6.2. The amplified hydrazone can thus act as a substrate for SjGST and binds in the targeted H-site.

The amplified hydrazones were re-synthesized and assayed against both GSTs and found to be inhibitors of GSH conjugation of CDNB, but poor solubility prevented the determination of accurate IC_{50} values at the higher concentrations necessary to assay weak binding compounds. To solve this problem, and simultaneously increase the potency of our DCL components, we conjugated GSH to aldehyde **1** using an S_NAr reaction. We anticipated that the highly soluble GSH tripeptide motif would act as an 'anchor' at the G-site, enabling exploration of the H-site with assorted hydrazide fragments. This approach, in which a known enzyme–substrate interaction is used for inhibitor discovery, is well exemplified in classical medicinal chemistry drug design, GST inhibition[51] and DCC methods. The IC_{50} value for SjGST inhibition of the anchored fragment **4** was measured in the CDNB conjugation assay as 280 μM.

Initial DCC experiments using GS-conjugated aldehyde **4** and the same ten hydrazides used previously confirmed the utility of aniline as a nucleophilic catalyst (Fig. 5). Equilibration was complete in 6 h, compared to 4 days in the absence of aniline, and each of the ten acylhydrazones were clearly identified by liquid chromatography-mass spectrometry (LC-MS) (see Supplementary Information). As before, clear amplifications could be observed for both GST

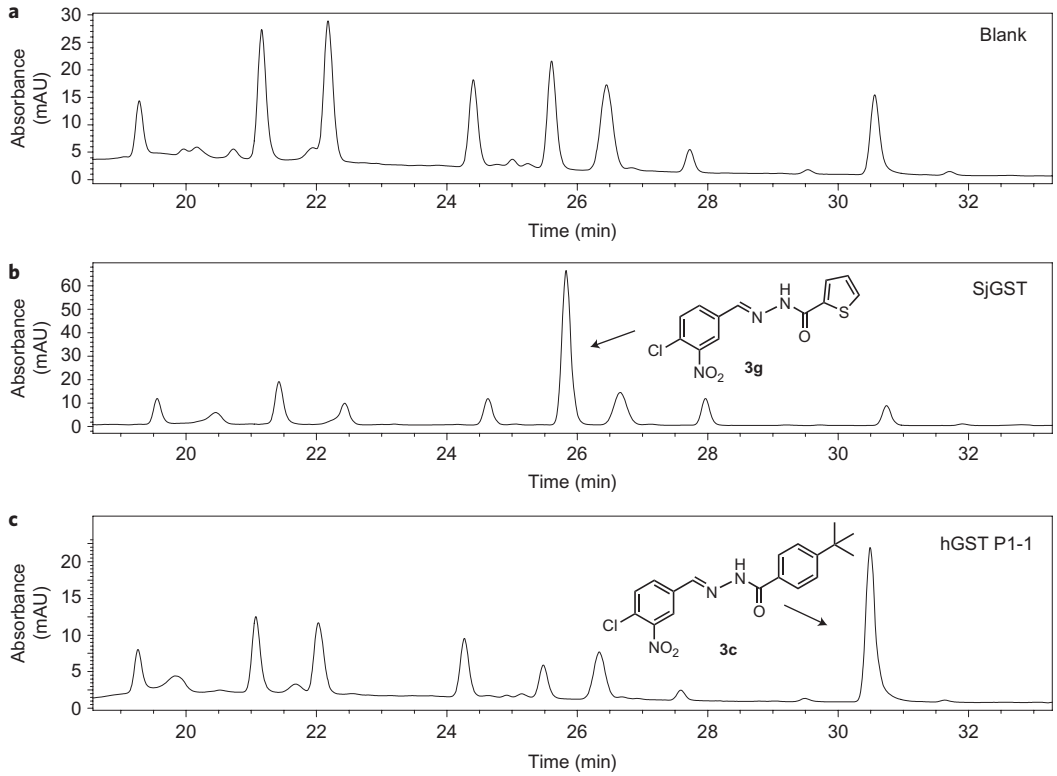


Figure 4 | GST-templated DCLs. a, DCL hydrazone composition in the absence of any target (blank). **b**, When the DCL is constituted in the presence of SjGST, the thiophene hydrazone **3g** is clearly amplified. **c**, Changing the target protein to hGSTP1-1 produces a different distrubution, in which the *t*-butylphenyl derivative **3c** is amplified. Targeted DCL conditions: GST (1 equiv.), aldehyde (5 μM), hydrazides (20 μM) and aniline (10 mM) in NH_4OAc buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h.

targets: in each case the same hydrazide fragment was selected as the best binder, thiophene (**5g**) for SjGST and *t*-butylphenyl (**5c**). Both components were amplified to over 300% of their concentrations in the blank DCL, at the expense of nearly all other competing hydrazones. Also of note, the anisyl sulfonylhydrazone **5j** underwent ~100% amplification using hGST P1-1 as the only other positively selected component. The most significant reductions in equilibrium concentrations occurred for **5b**, **f** and **i** (SjGST) and **5f**, **g** and **i** (hGST P1-1).

The GST-directed DCLs were synthesized with the protein present from the beginning of the experiment, that is, in the presence of aldehyde **4** and the ten hydrazides **2a**–**2j**. To verify that the amplification results were not due to a kinetic selection by means of target-accelerated synthesis, we added SjGST to the pre-equilibrated DCL. The same equilibrium distribution was achieved as is shown in Fig. 5, with hydrazone **5g** strongly amplified, indicating that the amplified components are the result of genuine thermodynamic selection. Further controls involved a BSA control experiment, which was negative, and DCL synthesis in the presence of a large excess of the non-selective GST inhibitor ethacrynic acid. Component amplification was completely suppressed for both SjGST and hGST P1-1 DCLs, indicating that the GST active site is saturated by the ethacrynic acid and cannot influence the DCL equilibrium composition.

We completed our protein-directed DCL studies by preparing a catalytically inactive SjGST mutant. It was of interest to see whether a functionally disabled enzyme would exert the same control and selectivity on DCL composition as the wild-type enzyme. The conserved Tyr 7 active site residue is known to play a critical role in GSH conjugation for the Sj class of GSTs, stabilizing the GSH thiolate anion through H-bonding from the phenol group, with enzymes lacking this residue being catalytically inactive[52]. We prepared a Y7F mutant of SjGST, in which the crucial tyrosine residue is replaced with phenylalanine. We observed essentially zero activity with this mutant in CDNB conjugation when compared with the wild-type SjGST. However, SjGST Y7F proved equally effective in controlling DCL composition, showing a clear preference for the same thiophene derivative **5g** as was amplified by the wild-type SjGST (see Supplementary Information).

Biological assay was then performed to establish whether the best binding compounds in the GST-directed DCLs were also the best inhibitors of the GST enzyme. To fully explore the isozyme-specific amplification effects of the two DCLs, we separately synthesized hydrazone conjugates **5a**–**5j** for study. We first confirmed that the amplified ligands **5c** and **5g** bound to SjGST and hGST P1-1 by isothermal calorimetry (ITC) (see Supplementary Information). We then studied their inhibitory activity towards SjGST and hGST P1-1 using the CDNB conjugation assay. The IC_{50} values were slightly higher for all hydrazones against hGST P1-1 compared to SjGST (data ranging from 59 to 126 μM and 22 to 63 μM, respectively; see Supplementary Information). For each isozyme, the DCC amplified hydrazone was the most active; thiophene **5g** had the lowest IC_{50} value (22 μM) among all the library members against SjGST, and *t*-butylphenyl **5c** had the lowest value among the four conjugates tested against hGST P1-1 (57 μM). The DCL hydrazone selection process has successfully extended inhibitor structure in the GST H-site, increasing potencies by sixfold for hGST P1-1 (331 to 57 μM) and by over tenfold for SjGST (279 to 22 μM) relative to the starting anchored aldehyde **4**.

Steady-state kinetic studies on the two amplified DCL components **5c** (hGST P1-1) and **5g** (SjGST) confirmed the expected competitive inhibition profile, with both compounds binding to the GST active sites. It was interesting to note slightly higher K_i values for both compounds when assayed against CDNB, a substrate for the H-site of the enzyme, relative to the endogenous G-site

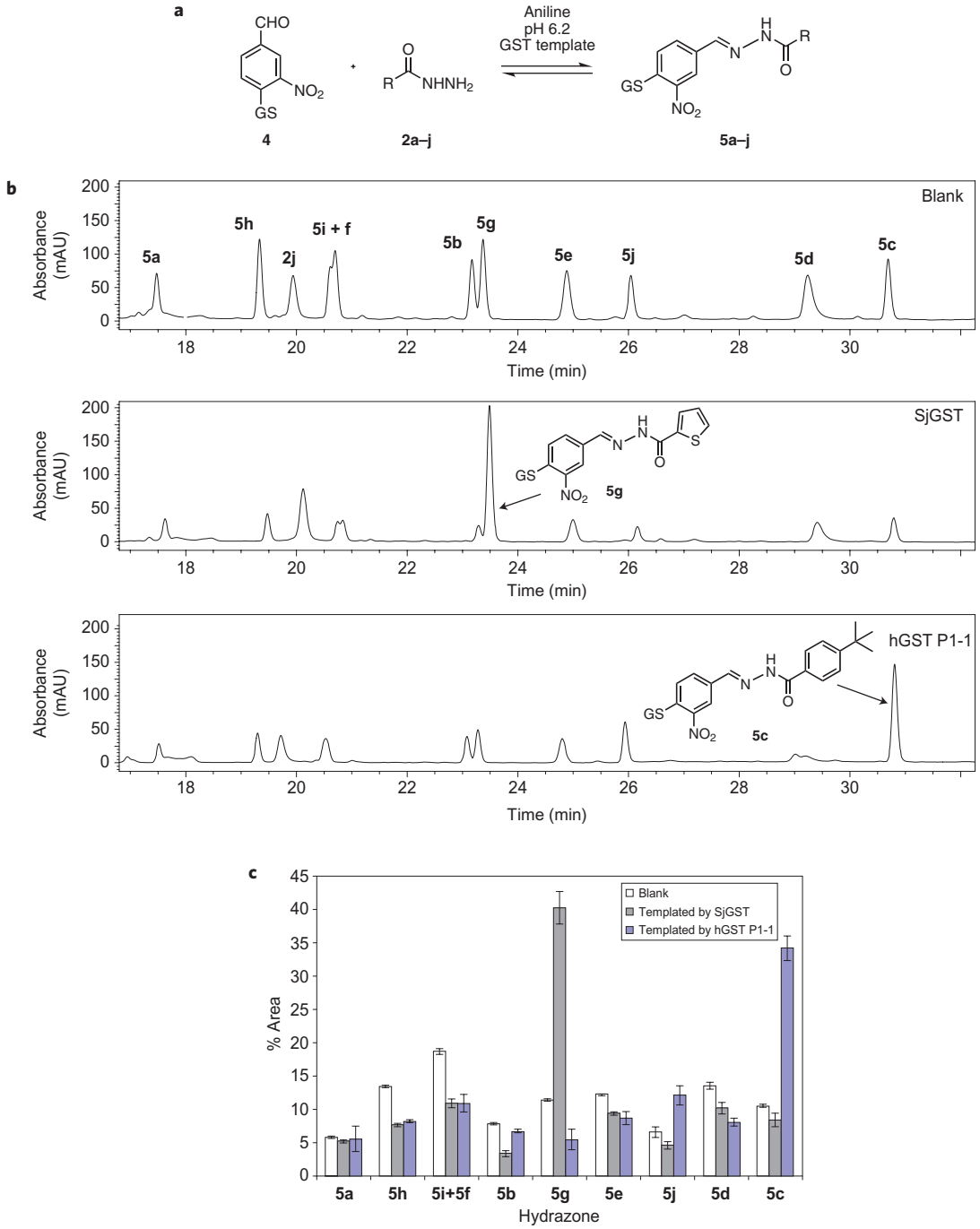


Figure 5 | GST-templated DCLs of GSH conjugates. a, Acyl hydrazone DCL based on GSH-conjugated aldehyde **4** (GS = *S*-linked glutathione). **b**, DCL hydrazone composition in the absence of target (blank), in the presence of SjGST and in the presence of hGSTP1-1. DCL conditions: GST (1 equiv.), aldehyde (5 μM), hydrazides (20 μM) and aniline (10 mM) in NH_4OAc buffer (50 mM, pH = 6.2) containing 15% DMSO for 16 h. **c**, Changes in DCL component concentration for blank, SjGST and hGST P1-1 DCLs. The error bars represent the standard deviation over three experiments.

ligand GSH. The affinity of the two hydrazone conjugates towards both GST G-sites was relatively close (data ranging from 5.25 to 7.19 μM), as would be expected for two compounds sharing a common GSH-tagged nitrobenzene fragment.

To obtain some molecular insight into the selectivity of our isozymes towards the two hydrazone inhibitors **5c** and **g**, we carried out a molecular modelling study. We surveyed the available GST structures in the protein data bank (PDB) and retrieved those that contained a bound GSH-based ligand. The binding sites of these structures, together with the bound ligands, were aligned, and it became evident that the glutathione portions overlaid well, being bound in very similar conformations in the G-sites (Fig. 6a). In contrast, the conjugate parts of the various ligands showed great diversity in their conformations within the H-site, an unsurprising result given the respective functions of the G- and H-sites. Detailed analysis of the superimposed crystal structures identified the GSH conjugate of 1,2-epoxy-3-(*p*-nitrophenoxy)propane (EPNP) (**6**) bound to cGST M1-1 (PDB code 1c72)[53] as the ligand that projected functionality into the H-site with the most similar geometry to the energy-minimized structure of hydrazone **5g** (Fig. 6b).

Analysis of the GST–EPNP complex shows the EPNP moiety orienting towards R107 and Q165 in the H-site of the enzyme

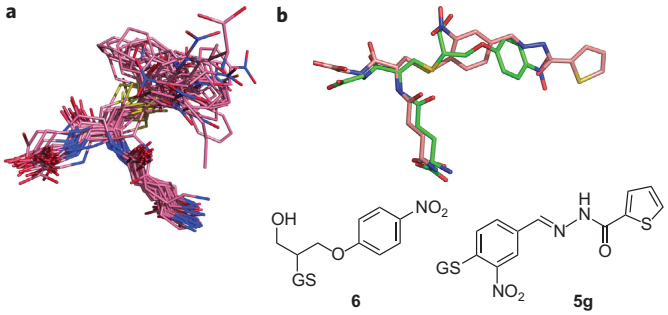


Figure 6 | GST ligands. a, Superposition of a selection of GST ligands from the PDB. **b**, Conformation of the GST-bound EPNP ligand **6** as found in the crystal structure of cGST M1-1 (PDB code 1c72, green carbon atoms), relative to the energy-minimized structure of compound **5g** (pink carbon atoms).

(Fig. 7c). The side chains of R107, F110, Q165, Q166 and F208 define the pocket that confines the EPNP moiety. On this basis, we could generate a binding model for SjGST with thiophene hydrazone **5g** and for hGST P1-1 with *t*-butyl hydrazone **5c** (Fig. 7). The interactions in the generated binding modes for SjGST in complex with **5g** (Fig. 7a) and for hGST P1-1 in complex with **5c** (Fig. 7b) between the glutathione moiety and the proteins are identical to those reported in previous publications[54,55]. We predict that the hydrazone group of **5g** forms hydrogen bonds to R103 and Q204 in SjGST, and equivalent interactions are observed for **5c** in complex with hGST P1-1. Residue V161 in SjGST and I161 in hGST P1-1 make hydrophobic interactions in our models with the ligands **5c** and **5g**.

As expected, the sub-pockets of both isoforms accommodating the hydrazones are rather hydrophobic, and complement the hydrophobic hydrazones amplified from the DCL. The thiophene hydrazone fits easily in the SjGST binding pocket, with only minor side-chain adjustments necessary (root mean square deviation (RMSD) 0.3 Å between model and crystal structure template), whereas the *t*-butylphenyl group would lead to a steric clash and would require some degree of induced fit in order to bind. Induced fit is also required to accommodate this ligand in the hGST P1-1 pocket, but in that case the binding mode could be stabilized by additional lipophilic interactions of the *t*-butyl group with Y103, H162 and I161. It is worth noting that Chern and colleagues have reported that mutations in that region of the H-site had a great impact on EPNP binding as a substrate, with mutation of cGST M1-1 Q165 to leucine (V161 in SjGST and I161 in hGST P1-1) reducing k_{cat}^{EPNP} by 59%, although K_m showed only small changes[53]. Because the amino acids in the equivalent pocket of SjGST and hGST P1-1 are not highly conserved, these residues have such a great influence on ligand binding that it is likely that these amino-acid exchanges across the isoforms are critical in determining ligand selectivity.

Conclusions

We have demonstrated that reversible synthesis of acylhydrazones can be compatible with protein targets by using aniline as a nucleophilic catalyst. The many advantages of this DCC tool (ready availability of easily customized building blocks, good kinetic and thermodynamic properties leading to ease of analysis, good biological compatibility forming amide-like linkages) may now be realized with biological targets. Most importantly, the acylhydrazone DCLs are truly adaptive, allowing amplification effects to be simply and directly related to structures present at equilibrium.

The GST enzyme proved extremely effective as a DCL template, with two isozymes from the GST family smoothly integrating with the small molecule assemblies and strongly amplifying the best binding components. The selected hydrazones showed increased inhibitory activity of over one order of magnitude from the starting GSH-tagged benzaldehyde **4**, validating the approach in the context of protein–ligand discovery. Interestingly, a single, small DCL composed of only ten members displays isozyme selectivity according to which variant of the GST enzyme is used as the template.

The study at hand has been deliberately confined to a small number of DCL components so as to thoroughly characterize equilibrium distributions and quantify amplifications with the aniline-catalysed hydrazone method. In principle, much larger hydrazone DCLs may be accessed to thoroughly explore chemical space, both within the GST H-site and for other biological targets[9,15]. It may not be possible, or even desirable, to accurately characterize the equilibrium distribution of such complex DCLs, but this will not be necessary if one simply seeks to identify prominently amplified components from a ligand discovery perspective.

To gain insight into isoform selectivity, we found that each amplified molecule could be effectively docked into its respective GST H-site, although the fine structural features of the SjGST versus hGST P1-1 H-site that discriminate between thiophene hydrazone **5g** and *t*-butylphenyl hydrazone **5c** are unclear at the present time. Structural determination of the complexes of various GST:GS–hydrazone conjugates will be needed for a deeper understanding of the factors that control H-site selectivity. Work in this

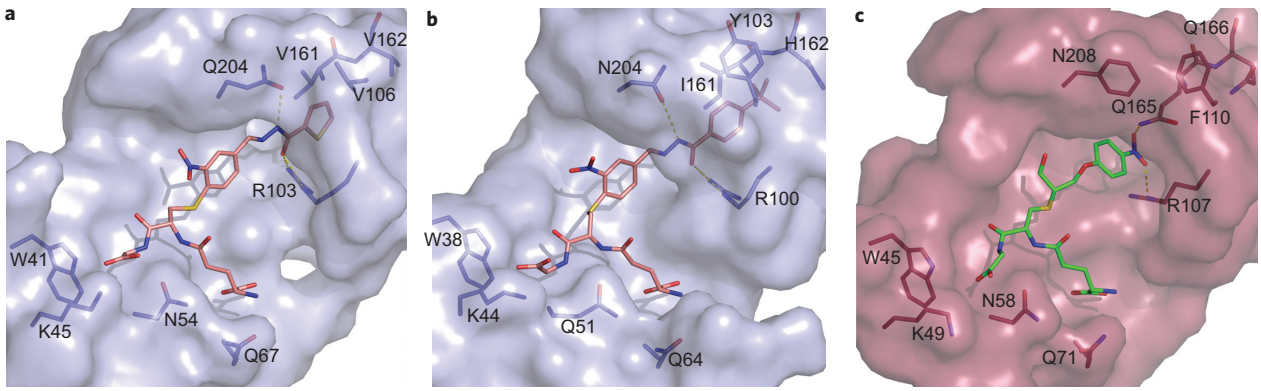


Figure 7 | Molecular modelling of amplified DCL components with the GST active site. a, Model of **5g** bound to SjGST. **b**, Model of **5c** bound to hGST P1-1. The binding pocket surfaces are shown in light blue and key amino acids as blue sticks. The ligands are represented in salmon pink, with atoms coloured by type. Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines. **c**, The EPNP–cGST M1-1 crystal structure (PDB code 1c72). The binding pocket surface is shown in raspberry pink and key amino acids as red sticks. The ligand is represented in green, with atoms coloured by type. Hydrogen bonds of the conjugated ligand parts are shown as yellow dotted lines.

area, together with applications of acylhydrazone DCC to other biological targets, is the subject of our current research.

Methods

Aniline catalysis of reversible hydrazone formation. The ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **1** (2 μl, 10 mM, DMSO) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (93 μl) and ammonium acetate buffer (845 μl, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature with occasional shaking, and was monitored periodically by HPLC to establish the blank composition until the relative populations of the hydrazones became constant. The pH of all samples was raised to 8 by the addition of NaOH (15 μl, 1 M, aqueous). LC-MS verified that each of the expected hydrazones was present in the DCL (HPLC conditions: column, Luna 5 μ C18(2), 30 mm × 4.6 mm, and Luna 5 μ C18(2), 50 mm × 4.6 mm, in sequence; flow rate, 1 ml min^{-1}; wavelength, 254 nm; temperature, 23 °C; gradient H_2O/MeCN (0.01% TFA) from 95% to 80% over 6 min, then to 45% over 30 min, and eventually to 5% over 5 min) (Fig. 2d). The DCL was then re-synthesized in the absence of aniline, and the HPLC traces at different time intervals were compared (Fig. 2c).

Templated DCL aldehyde 1. SjGST (111 μl, 180 μM, in potassium phosphate buffer 0.1 M, pH 6.8), the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **1** (2 μl, 10 mM, DMSO) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (93 μl) and ammonium acetate buffer (734 μl, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature, with occasional shaking, for 12 h. The pH of the sample was raised to 8 by the addition of NaOH (15 μl, 1 M, aqueous), and the protein was removed by ultrafiltration using a 10,000 MWCO filter (Vivaspin). HPLC analysis was performed and the traces were compared with the blank composition (HPLC conditions: column, Luna 5 μ C18(2), 30 mm × 4.6 mm, and Luna 5 μ C18(2), 50 mm × 4.6 mm, in sequence; flow rate, 1 ml min^{-1}; wavelength, 254 nm; temperature, 23 °C; gradient H_2O/MeCN (0.01% TFA) from 95% to 80% over 6 min, then to 45% over 30 min, and eventually to 5% over 5 min).

DCL composition was identical, regardless of whether the SjGST was present from the beginning or added after pre-equilibration, but equilibration took more than 24 h in the latter case.

For the hGST P1-1 templated library, the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **1** (2 μl, 10 mM, DMSO), aniline (10 μl, 1 M, DMSO) and hGST P1-1 (100 μl, 200 μM, in potassium phosphate buffer 0.1 M, pH 6.8) were added to a mixture of DMSO (93 μl) and ammonium acetate buffer (734 μl, 50 mM, pH 6.2). After equilibration for 12 h, the DCL was analysed using HPLC. Control experiments were performed using the same equivalents of BSA in place of GST.

Conjugate DCLs. To establish the blank DCL composition, the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **5** (5 μl, 10 mM, aqueous) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (96 μl) and ammonium acetate buffer (839 μl, 50 mM, pH 6.2). The DCL was allowed to stand at room temperature, with occasional shaking, and was monitored periodically by HPLC to establish the blank composition until the relative populations of the hydrazones became constant. The pH of all the samples was increased to 8 by the addition of NaOH (15 μl, 1 M, aqueous). LC-MS verified that each of the expected hydrazones was present in the DCL (Fig. 5) (HPLC conditions: column, Luna 5 μ C18(2), 50 mm × 4.6 mm, and Luna 5 μ C18(2), 250 mm × 4.6 mm, in sequence; flow rate, 1 ml min^{-1}; wavelength, 254 nm; temperature, 23 °C; gradient H_2O/MeCN (0.01% TFA) from 95% to 5% over 40 min).

For re-synthesizing the DCL in the presence of the protein SjGST (278 μl, 180 μM, in potassium phosphate buffer 0.1 M, pH 6.8), the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **5** (5 μl, 10 mM, DMSO) and aniline (10 μl, 1 M, DMSO) were added to a mixture of DMSO (96 μl) and ammonium acetate buffer (561 μl, 50 mM, pH 6.2). The DCL templated by hGST P1-1 was synthesized by adding the ten hydrazides **2a–j** (10 × 5 μl, 10 mM, DMSO), aldehyde **5** (5 μl, 10 mM, DMSO), aniline (10 μl, 1 M, DMSO) and hGST P1-1 (250 μl, 200 μM, in potassium phosphate buffer 0.1 M, pH 6.8) to a mixture of DMSO (96 μl) and ammonium acetate buffer (589 μl, 50 mM, pH 6.2). The DCLs were allowed to stand at room temperature for 12 h, after which the pH was raised to 8 by the addition of NaOH (15 μl, 1 M). The protein was filtered off using a centrifuge filter of MWCO 10,000 followed by analysis of the filtrate by HPLC using conditions similar to those listed above.

Molecular modelling. To establish ligand alignment, the superposition of GST ligands was carried out using Relibase + 3.0.0 (ref. 56). A search was first performed to find binding sites that share a sequence identity between 40 and 100% with the target GST crystal structure 1m9a. The resulting 38 structures with bound ligand were superimposed by using binding site residues only. Finally, the ligands from the superimposed structures were extracted and visually analysed.

To carry out a binding mode prediction with Moloc[57], the SjGST crystal structure (PDB code 1m9a SjGST – *S*-hexyl–GSH complex) and the hGST P1-1–GSH complex crystal structure (PDB code 6gss)[54] were used as starting conformations for binding mode generation. The glutathione groups of the synthesized ligands were mapped onto the glutathione groups of the ligands bound to the crystal structures. The hydrophobic hydrazone groups of the synthesized ligands were oriented towards the cavity, lying at the end of the *S*-hexyl site, as observed for the EPNP ligand bound to cGSTM1-1 (PDB code 1c72). In the next step, the protein in complex with the modelled ligand was minimized, considering the ligand as fully flexible. For the protein all residues were kept rigid, except for the amino acids that define the pocket at the end of the *S*-hexyl site (R103, V106, V161, V162, Q204 for SjGST and R100, Y103, I161, H162, N204 for hGST P1-1).

Received 7 December 2009; accepted 30 March 2010; published online 16 May 2010

References

1. Lehn, J.-M. Dynamic combinatorial chemistry and virtual combinatorial libraries. *Chem. Eur. J.* **5,** 2455–2463 (1999).
2. Rowan, S. J., Cantrill, S. J., Cousins, G. R. L., Sanders, J. K. M. & Stoddart, J. F. Dynamic covalent chemistry. *Angew. Chem. Int. Ed.* **41,** 898–952 (2002).
3. Corbett, P. T. *et al.* Dynamic combinatorial chemistry. *Chem. Rev.* **6,** 3652–3711 (2006).
4. Ladame, S. Dynamic combinatorial chemistry: on the road to fulfilling the promise. *Org. Biomol. Chem.* **6,** 219–226 (2008).
5. Erlanson, D. A. *et al.* Site-directed ligand discovery. *Proc. Natl Acad. Sci. USA* **97,** 9367–9372 (2000).
6. Corbett, A. R., Cheeseman, J. D., Kazlauskas, R. J. & Gleason, J. L. Pseudodynamic combinatorial libraries: a receptor-assisted approach for drug discovery. *Angew. Chem. Int. Ed.* **43,** 2432–2436 (2004).
7. Hochgürtel, M. *et al.* Target-induced formation of neuraminidase inhibitors from *in vitro* virtual combinatorial libraries. *Proc. Natl Acad. Sci. USA* **99,** 3382–3387 (2002).
8. Zameo, S., Vauzeilles, B. & Beau, J. M. Dynamic combinatorial chemistry: lysozyme selects an aromatic motif that mimics a carbohydrate residue. *Angew. Chem. Int. Ed.* **44,** 965–969 (2005).
9. McNaughton, B. R., Gareiss, P. C. & Miller, B. L. Identification of a selective small-molecule ligand for HIV-1 frameshift-inducing stem-loop RNA from an 11,325 member resin bound dynamic combinatorial library. *J. Am. Chem. Soc.* **129,** 11306–11307 (2007).
10. Gareiss, P. C. *et al.* Dynamic combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA-MBNL1 interaction *in vitro*: discovery of lead compounds targeting myotonic dystrophy (DM1). *J. Am. Chem. Soc.* **130,** 16254–16261 (2008).
11. Scott, D. E., Dawes, G. J., Ando, M., Abell, C. & Ciulli, A. A fragment-based approach to probing adenosine recognition sites by using dynamic combinatorial chemistry. *ChemBioChem* **10,** 2772–2779 (2009).
12. Eliseev, A. V. & Nelen, M. I. Use of molecular recognition to drive chemical evolution 1. Controlling the composition of an equilibrating mixture of simple arginine receptors. *J. Am. Chem. Soc.* **119,** 1147–1148 (1997).
13. Hioki, H. & Still, W. C. Chemical evolution: a model system that selects and amplifies a receptor for the tripeptide (D)Pro(L)Val(D)Val. *J. Org. Chem.* **63,** 904–905 (1998).
14. Wietor, J. L., Pantos, G. D. & Sanders, J. K. M. Templated amplification of an unexpected receptor for C-70. *Angew. Chem. Int. Ed.* **47,** 2689–2692 (2008).
15. Ludlow, R. F. & Otto, S. Two-vial LC-MS identification of ephedrine receptors from a solution phase dynamic combinatorial library of over 9,000 components. *J. Am. Chem. Soc.* **130,** 12218–12219 (2008).
16. Turega, S. M., Lorenz, C., Sadownik, J. W. & Philp, D. Target-driven selection in a dynamic nitrone library. *Chem. Commun.* 4076–4078 (2008).
17. Xu, S. & Giuseppone, N. Self-duplicating amplification in a dynamic combinatorial library. *J. Am. Chem. Soc.* **130,** 1826–1827 (2008).
18. Sadownik, J. W. & Philp, D. A simple synthetic replicator amplifies itself from a dynamic reagent pool. *Angew. Chem. Int. Ed.* **47,** 9965–9970 (2008).
19. Nguyen, R., Allouche, L., Buhler, E. & Giuseppone, N. Dynamic combinatorial evolution within self-replicating supramolecular assemblies. *Angew. Chem. Int. Ed.* **48,** 1093–1096 (2009).
20. Chichak, K. S. *et al.* Molecular Borromean rings. *Science* **304,** 1308–1312 (2004).
21. Lam, R. T. S. *et al.* Amplification of acetylcholine-binding catenanes from dynamic combinatorial libraries. *Science* **308,** 667–669 (2005).
22. Au-Yeung, H. Y., Pantos, G. D. & Sanders, J. K. M. Dynamic combinatorial synthesis of a catenane based on donor–acceptor interactions in water. *Proc. Natl Acad. Sci. USA* **106,** 10466–10470 (2009).
23. Tauk, L., Schroder, A. P., Decher, G. & Giuseppone, N. Hierarchical functional gradients of pH-responsive self-assembled monolayers using dynamic covalent chemistry on surfaces. *Nature Chem.* **1,** 649–656 (2009).
24. Fujii, S. & Lehn, J.-M. Structural and functional evolution of a library of constitutional dynamic polymers driven by alkali metal ion recognition. *Angew. Chem. Int. Ed.* **48,** 7635–7638 (2009).
25. Kindermann, M., Stahl, I., Reimold, M., Pankau, W. M. & von Kiedrowski, G. Systems chemistry: kinetic and computational analysis of a nearly exponential organic replicator. *Angew. Chem. Int. Ed.* **44,** 6750–6755 (2005).
26. Ludlow, R. F. & Otto, S. Systems chemistry. *Chem. Soc. Rev.* **37,** 101–108 (2008).
27. Shi, B., Stevenson, R., Campopiano, D. J. & Greaney, M. F. Discovery of glutathione S-transferase inhibitors using dynamic combinatorial chemistry. *J. Am. Chem. Soc.* **128,** 8459–8467 (2006).

28. Ramstrom, O. & Lehn, J.-M. *In situ* generation and screening of a dynamic combinatorial carbohydrate library against concanavalin A. *ChemBioChem* **1,** 41–48 (2000).
29. Otto, S., Furlan, R. L. E. & Sanders, J. K. M. Dynamic combinatorial libraries of macrocyclic disulfides in water. *J. Am. Chem. Soc.* **122,** 12063–12064 (2000).
30. Nicolaou, K. C. *et al.* Target-accelerated combinatorial synthesis and discovery of highly potent antibiotics effective against vancomycin-resistant bacteria. *Angew. Chem. Int. Ed.* **39,** 3823–3828 (2000).
31. Milanesi, L., Hunter, C. A., Sedelnikova, S. E. & Waltho, J. P. Amplification of bifunctional ligands for calmodulin from a dynamic combinatorial library. *Chem. Eur. J.* **12,** 1081–1087 (2006)
32. Shi, B. & Greaney, M. F. Reversible Michael addition of thiols as a new tool for dynamic combinatorial chemistry. *Chem. Commun.* 886–888 (2005).
33. Huc, I. & Lehn, J.-M. Virtual combinatorial libraries: dynamic generation of molecular and supramolecular diversity by self-assembly. *Proc. Natl Acad. Sci. USA* **94,** 2106–2110 (1997).
34. Cousins, G. R. L., Poulsen, S. A. & Sanders, J. K. M. Dynamic combinatorial libraries of pseudo-peptide hydrazone macrocycles. *Chem. Commun.* 1575–1576 (1999).
35. Furlan, R. L. E., Ng, Y. F., Otto, S. & Sanders, J. K. M. A new cyclic pseudopeptide receptor for Li^+ from a dynamic combinatorial library. *J. Am. Chem. Soc.* **123,** 8876–8877 (2001).
36. Roberts, S. L., Furlan, R. L. E., Cousins, G. R. L. & Sanders, J. K. M. Simultaneous selection, amplification and isolation of a pseudo-peptide receptor by an immobilised N-methyl ammonium ion template. *Chem. Commun.* 938–939 (2002).
37. Liu, J. Y., West, K. R., Bondy, C. R. & Sanders, J. K. M. Dynamic combinatorial libraries of hydrazone-linked pseudo-peptides: dependence of diversity on building block structure and chirality. *Org. Biomol. Chem* **5,** 778–786 (2007).
38. Bunyapaiboonsri, T. *et al.* Dynamic deconvolution of a pre-equilibrated dynamic combinatorial library of acetylcholinesterase inhibitors. *ChemBioChem* **2,** 438–444 (2001).
39. Bunyapaiboonsri, T., Ramstrom, H., Ramstrom, O., Haiech, J. & Lehn, J.-M. Generation of bis-cationic heterocyclic inhibitors of *Bacillus subtilis* HPr kinase/phosphatase from a ditopic dynamic combinatorial library. *J. Med. Chem.* **46,** 5803–5811 (2003).
40. Poulsen, S. A. Direct screening of a dynamic combinatorial library using mass spectrometry. *J. Am. Soc. Mass Spectrom.* **17,** 1074–1080 (2006).
41. Cordes, E. H. & Jencks, W. P. Nucleophilic catalysis of semicarbazone formation by anilines. *J. Am. Chem. Soc.* **84,** 826–831 (1962).
42. Dirksen, A., Dirksen, S., Hackeng, T. M. & Dawson, P. E. Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry. *J. Am. Chem. Soc.* **128,** 15602–15603 (2006).
43. Dirksen, A. & Dawson, P. E. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. *Bioconjugate Chem.* **19,** 2543–2548 (2008).
44. Rodriguez-Docampo, Z. & Otto, S. Orthogonal or simultaneous use of disulfide and hydrazone exchange in dynamic covalent chemistry in aqueous solution. *Chem. Commun.* 5301–5303 (2008).
45. Hayes, J. D., Flanagan, J. U. & Jowsey, I. R. Glutathione transferases. *Annu. Rev. Pharmacol. Toxicol.* **45,** 51–88 (2005).
46. Mahajan, S. & Atkins, W. M. The chemistry and biology of inhibitors and pro-drugs targeted to glutathione S-transferases. *Cell. Mol. Life Sci.* **62,** 1221–1233 (2005).
47. Li, W.-S. *et al.* Overcoming the drug resistance in breast cancer cells by rational design of efficient glutathione S-transferase inhibitors. *Org. Lett.* **12,** 20–23 (2010).
48. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nature Chem.* **1,** 187–192 (2009).
49. Jao, S. C., Chen, J., Yang, K. & Li, W. S. Design of potent inhibitors for *Schistosoma japonica* glutathione S-transferase. *Bioorg. Med. Chem.* **14,** 304–318 (2006).
50. Tew, K. D. Glutathione-associated enzymes in anticancer drug-resistance. *Cancer Res.* **54,** 4313–4320 (1994).
51. Lyon, R. P., Hill, J. J. & Atkins, W. M. Novel class of bivalent glutathione S-transferase inhibitors. *Biochemistry* **42,** 10418–10428 (2003).
52. Andújar-Sánchez, M. *et al.* Crystallographic and thermodynamic analysis of the binding of S-octylglutathione to the Tyr 7 to Phe mutant of glutathione S-transferase from *Schistosoma japonicum*. *Biochemistry* **44,** 1174–1183 (2005).
53. Chern, M. K. *et al.* Tyr115, Gln165 and Trp209 contribute to the 1,2-epoxy-3-(p-nitrophenoxy)propane-conjugating activity of glutathione S-transferase cGSTM1-1. *J. Mol. Biol.* **300,** 1257–1269 (2000).
54. Oakley, A. J. *et al.* The structures of human glutathione transferase P1-1 in complex with glutathione and various inhibitors at high resolution. *J. Mol. Biol.* **274,** 84–100 (1997).
55. Cardoso, R. M. F., Daniels, D. S., Bruns, C. M. & Tainer, J. A. Characterization of the electrophile binding site and substrate binding mode of the 26-kDa glutathione S-transferase from *Schistosoma japonicum*. *Proteins* **51,** 137–146 (2003).
56. Bergner, A., Gunther, J., Hendlich, M., Klebe, G. & Verdonk, M. Use of relibase for retrieving complex three-dimensional interaction patterns including crystallographic packing effects. *Biopolymers* **61,** 99–110 (2001).
57. Gerber, P. R. & Muller, K. MAB, a generally applicable molecular-force field for structure modeling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **9,** 251–268 (1995).

Acknowledgements

The authors would like to thank EastChem for the award of a studentship to V.T.B. and the Marie Curie Early Stage Training Network (Syn4chembio) and School of Chemistry at Edinburgh for awarding a studentship to A.M.C. R.B. is supported by an EC Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 223461. M.F.G. is an Engineering and Physical Sciences Research Council (EPSRC) Leadership Fellow. The authors thank A. Cooper (University of Glasgow) for ITC measurements and helpful discussions. N. Petitjean is thanked for the synthesis of hydrazone–GSH conjugates.

Author contributions

V.T.B., A.M.C., D.J.C. and M.F.G. conceived and designed the experiments, V.T.B. and A.M.C. performed the experiments, and T.L., R.B. and A.M.C. carried out molecular modelling. All authors discussed the results and co-wrote the manuscript.

Additional information

The authors declare no competing financial interests. Supplementary information and chemical compound information accompany this paper at www.nature.com/naturechemistry. Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/. Correspondence and requests for materials should be addressed to D.J.C. and M.F.G.

Nitrilotriacetic Acid-Derivatized Quantum Dots for Simple Purification and Site-Selective Fluorescent Labeling of Active Proteins in a Single Step

Manish Gupta, Anne Caniard, Ángeles Touceda-Varela, Dominic J. Campopiano, and Juan C. Mareque-Rivas*

Received July 3, 2008; Revised Manuscript Received August 16, 2008

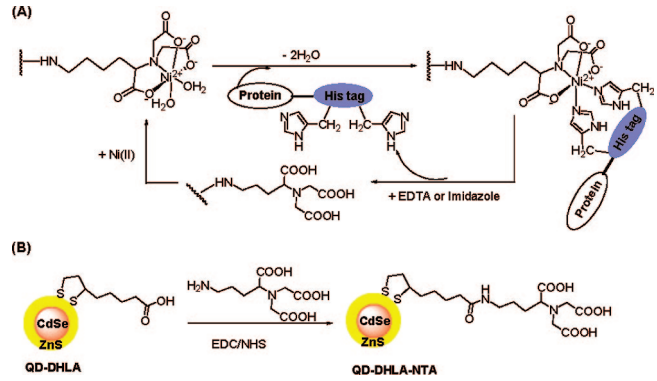
We demonstrate that QDs coated with nitrilotriacetic acid (NTA) bound to Ni^{2+} can be used to reversibly and selectively bind, purify, and fluorescently label His_6-tagged (N-terminal) glutathione S-transferase (GST) in one step with retention of enzymatic activity. We find binding to be less effective in the absence of the His_6-tag or Ni^{2+} ions.

Colloidal semiconductor nanocrystals (quantum dots, QDs) have emerged as powerful fluorescent probes for biological imaging applications (*1*−*5*). QDs have several advantages over small organic dyes and fluorescent proteins such as size-tuneable photoluminescence, wide excitation−narrow emission properties, improved brightness, and high resistance to photobleaching and degradation. To conjugate QDs to biomolecules, the QD surface is derivatized in such a way as to allow the attachment of the biomolecule through a covalent bond, electrostatic, or hydrophobic interactions. A typical biomolecule contains many residues capable of forming covalent and noncovalent linkages with the QD. The ability to control the site of attachment is important to ensure that the biomolecule bound to the QD is still active. Site-specific noncovalent binding of QDs to biomolecules has been achieved by exploiting carbohydrate−lectin and streptavidin−biotin interactions (*6, 7*).

A common approach to facilitate protein purification involves the use of genetically encoded oligohistidine (His_n) tags (*8*−*10*). Since His_n-tags are recognized by nitrilotriacetic acid (NTA) complexes of nickel(II) (Scheme 1), purification of these proteins is achieved by passing the protein mixture through chromatography columns containing Ni-NTA resins (*11*−*13*). Given the widespread use of His_n-tags and the fact that they can be introduced into regions of peptides where they do not disturb protein structure and function such as at the termini or in loops, they provide a convenient way to achieve site-specific binding for many applications (*14*−*18*). Hence, the fluorescent labeling of specific proteins by encoding His_n-tags on them coupled to the recognition of these tags by Ni-NTA has become popular. Initially, this was done using organic dye−NTA conjugates (*19, 20*), but very recently it has been extended to QDs. Thus, during the final stages of preparation of this manuscript the first study showing that Ni-NTA-containing QDs that can be applied to imaging His_6-tagged proteins in live cells has appeared (*21*). This is interesting because it has been shown that His_n-tags can also bind with very high affinity ($K_d \approx 1$ nM) to QDs with carboxylic acid functionalities requiring fewer synthetic steps and cheaper reagents (*22*). In another study, however, it was reported that QDs with carboxylic acids presented at the surface bind His-tagged proteins only in the presence of Ni^{2+} cations (*23*). Thus, it is interesting to investigate the advantages, disadvantages and applicability of each of these functionalities for the formation of QD−protein conjugates.

Here, we investigate binding of His_6-tagged and untagged GST as model protein to CdSe-ZnS core−shell nanoparticles with carboxylates, NTA, and Ni^{2+}-bound NTA at the surface.

Scheme 1[a]



[a] (A) Reversible binding between a His-tagged protein and Ni(NTA). (B) Synthesis of the QD functionalized with NTA.

QDs−GST conjugates are interesting because GSTs catalyze the nucleophilic addition of GSH to the electrophilic center of a range of nonpolar substrates as a way of detoxifying a wide range of harmful endogenous and xenobiotic compounds (*24*), and in drug resistance mechanisms (*25*). Since GST activity requires not only the formation of a dimeric structure (i.e., protein−protein interactions) but also binding of both GSH and an acceptor substrate (i.e., protein−substrate interactions), it is a good enzyme to investigate the effect of specific and nonspecific binding of QDs on enzymatic activity. We report that the combination of His_6-tagged GST with CdSe-ZnS core−shell nanoparticles coated with Ni^{2+}-bound NTA gives the best results. The resulting QD−protein conjugate is strongly fluorescent and readily purified by filtration or ultracentrifugation, which should enable widespread use of these QDs as a "two-in-one" purification−fluorescence labeling tool. Moreover, the QD binds the His_6-tagged GST with little disruption of its enzymatic activity, whereas QD binding is weaker and disrupts enzymatic activity when GST lacks a His-tag or when Ni^{2+} is unavailable. This is a significant result in that it shows that by using this simple construct protein purification, fluorescent tagging, and precise positioning of the fluorescent probe to preserve structural/functional properties are all accomplished in a single, inexpensive step. It shows also that QDs presenting Ni-NTA at the surface provide a good alternative to site-specifically binding His-tagged proteins if, like in this case, the easier-to-make QDs with carboxylic acids at the surface fail to do so (*vide infra*).

The QDs decorated with NTA were easily prepared from the reaction of CdSe-ZnS core−shell dihydrolipoic acid (DHLA)-

* Corresponding author. E-mail: juan.mareque@ed.ac.uk.

10.1021/bc800273j CCC: $40.75
Published on Web 09/19/2008

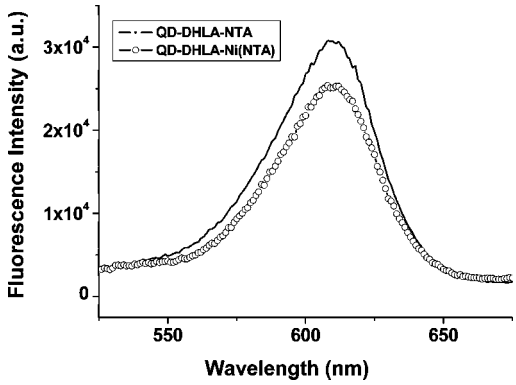


Figure 1. Photoluminescence spectra for CdSe-ZnS core−shell QDs coated with DHLA-NTA and DHLA-Ni(NTA) (spectra were acquired in 20 mM PBS pH 6.7, excitation at 350 nm; $T = 293$ K). The integrated emission decreases 15% upon Ni^{2+} addition.

capped QDs (*26*) with commercially available *N,N*-bis(carboxymethyl)-L-lysine hydrate using EDC and *N*-hydroxysuccinimide as coupling agents in phosphate buffer solution (Scheme 1). The product is purified by filtration with a Nanosep 100K centrifugal device (Pall Corporation). The composition of the QDs was examined by X-ray photoelectron spectroscopy (XPS). The XPS spectra showed the main diagnostic peaks of the product QDs: a 2s peak at 400.1 eV due to N, and 2p3 and 2p1 peaks at 857.1 and 874.1 eV, respectively, due to Ni^{2+} (Supporting Information).

The photoluminescence intensity of the Ni-NTA-capped QD is ca. 85% that of the NTA-capped QD (Figure 1). This is important because the paramagnetic Ni^{2+} was found to strongly quench the photoluminescence of some organic dyes, limiting their applications (*19*).

We selected GST from the helminth worm *Schistosoma japonica* (SjGST, 26 kDa monomer) as our target enzyme because it is amenable to recombinant overexpression in *E. coli* as a His_6-tagged construct and has been well-characterized (*9*). It is important to note that the His_6-tag was genetically fused at the N-terminus. The noncovalent attachment of His_6-tagged and untagged GST to the QD surface before and after derivatization with Ni-NTA was analyzed by SDS-PAGE. A solution of the QD or PBS (as control) was incubated with the corresponding enzyme for 2 h and passed through a Nanosep 300K centrifugal device. The retenate was redissolved in PBS buffer, and both retenate and filtrate were analyzed by SDS-PAGE. In the absence of the QD, His_6-tagged and untagged GST were found only in the filtrate. Several reports have shown that the DHLA-coated CdSe-ZnS core−shell QDs are capable of binding His-tagged proteins by coordination to Zn^{2+} ions at the nanocrystal surface (*22, 26*−*28*). Our SDS-PAGE studies, however, did not find protein in the retenate (Figure 2). Lack of binding could be due to steric hindrance at the N-terminus location of the His_6-tag preventing access to the Zn^{2+} atoms of the nanoparticle. It is also possible that the different synthetic procedures used to prepare QDs lead to subtle changes at the QD surface which affect binding of biomolecules. However, it is worth noting that it is not rare for His-tagged proteins to exhibit different properties depending on whether the His-tag is at the N- or C-terminus, and that in the studies reporting direct His-tag binding to carboxylate-coated QDs the His-tag was located at the C-terminus (*22, 26*−*28*). In contrast, using the same experimental conditions the Ni-NTA-capped QDs immobilized both enzymes. Moreover, we found more His_6-tagged than untagged GST in the retenate. Binding was also investigated in the presence of high salt concentrations. His_6-GST binding to Ni-NTA-capped QDs was not affected by 1 M NaCl. In contrast, untagged GST did not bind to the QDs under these conditions,

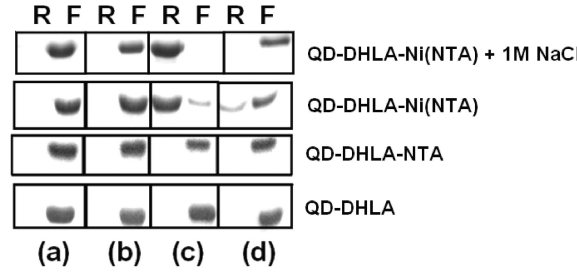


Figure 2. SDS-PAGE of the retenate (R) and filtrate (F) after ultrafiltration through a Nanosep 300K filter of His_6-GST (a), untagged GST (b), His_6-GST incubated with QD (c), and untagged GST incubated with QD (d). In each case, the enzyme and QD concentrations were 16.5 μM and 9.0 μM, respectively.

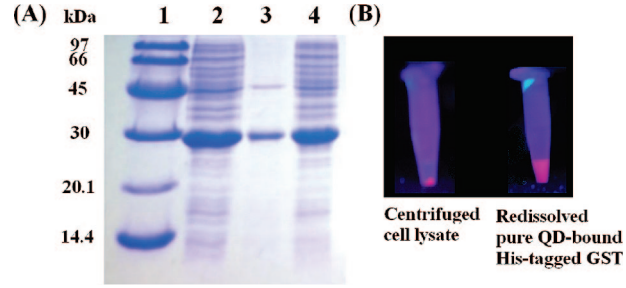


Figure 3. (A) SDS-PAGE studies of the cell lysate containing His_6-tagged GST (lane 2) and proteins released from the Ni-NTA-coated QDs treated with PBS containing 0.5 M imidazole (lane 3) and supernatant (lane 4) after ultracentrifugation. Lane 1 is the molecular weight marker. (B) Images of the cell lysate after ultracentrifugation and of the pure QD-bound His_6-tagged GST.

which suggests it is predominantly electrostatic. Thus, high salt concentrations can be used to avoid binding of untagged proteins while ensuring binding of the desired His_6-tagged target. The enzyme was easily released from the QD surface upon addition of 0.5 M imidazole, which competes for the Ni^{2+} binding sites. Thus, decorating the surface of the QD with Ni^{2+} complexes of NTA seems a good approach for noncovalent site-specific fluorescent labeling of proteins, which can be used for instance if carboxylate-functionalized QDs lacking Ni^{2+} ions fail. Potential advantages of attaching Ni-NTA units to QDs could be stronger interactions with the His-tag ($K_d \approx 10^{-13}$ M) (*18*) and less sensitivity to steric hindrance and surface properties by being further away from the nanocrystal surface.

Recently, the value of magnetic nanoparticles as affinity probes to selectively trap and separate His-tagged proteins from cell lysates has been elegantly demonstrated (*29*−*32*). The protein purification efficiency of the Ni-NTA-capped QDs was investigated by incubating cell lysates containing His_6-tagged GST for 2 h. Remarkably, pure fluorescently labeled GST was obtained simply by ultracentrifugaton of this mixture (Figure 3). Thus, by using the Ni-NTA-capped QDs it is possible to purify and fluorescently label His-tagged proteins in a single step. Current methods for efficiently purifying and fluorescently labeling His-tagged proteins need various labor-intensive and expensive steps, such as conjugation of NTA derivatives on support materials or the preparation of suitable magnetic nanoparticles for purification purposes, followed by the attachment of fluorescent tags. Another construct suitable for one-step protein purification and site-specific labeling was recently developed and involves organic fluorophore-doped Ni-NTA-modified silica nanoparticles (*33*).

In order to obtain information about the effect of QD binding on the catalytic activity of GST, we have used 1-chloro-2,4-dinitrobenzene (CDNB) as substrate. The GST-catalyzed reaction of GSH with CDNB produces a dinitrophenyl thioether

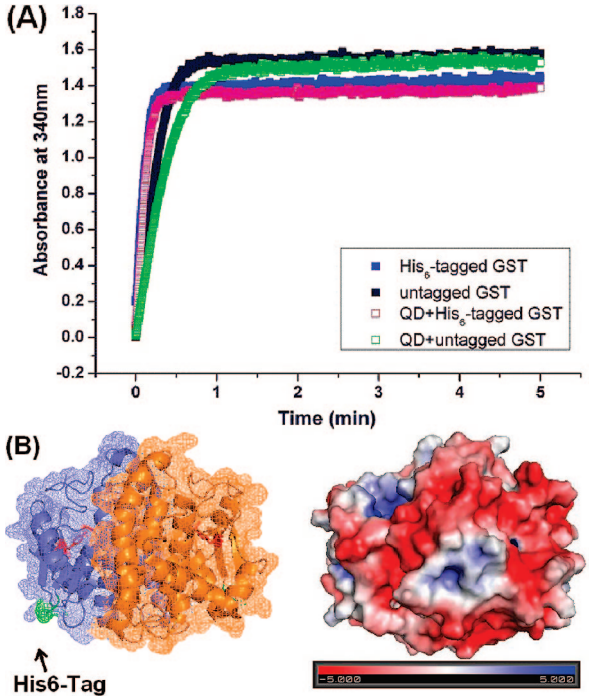


Figure 4. (A) Activity of His_6-tagged (N-terminal) and untagged GST in the presence of Ni-NTA coated QDs. Conditions: [GSH] = 8 mM, [CDNB] = 2 mM in PBS (pH 6.7), $T = 298$ K. The QD alone did not have any activity. (B) X-ray crystal structure of the SjGST homodimer highlighting the catalytically crucial Tyr7 residue in red, and the N-terminus site for the His_6-tag in green (left); surface charge distributions (right).

which can be conveniently detected spectrophotometrically at 340 nm (*34*). His_6-tagged GST and untagged GST were incubated with the same concentration of Ni-NTA-capped QD. We have found that His_6-tagged GST retains its activity after binding to the QD, whereas the untagged GST loses activity (Figure 4). We suggest that the ability of the His_6-tag to control the position of the Ni-NTA-capped QD relative to the GST active site is responsible for preserving the activity of the enzyme. The X-ray crystal structure of SjGST (*35*) shows that the N terminus, which is where the His_6 tag was placed, is ca. 25 Å away from the essential catalytic residue Tyr7 (Figure 4). We have examined the distribution of positively and negatively charged residues and found that there are positive and negative regions close to the active site. These are sites where in the absence of the His_6 tag nonspecific electrostatic binding could occur, disrupting the enzyme activity. By comparing the activity of the enzyme which did not bind to the QD with that of the enzyme before incubation with QD, we estimated the protein binding capacity and number of His_6-GST molecules immobilized on each QD (~16). This surface coverage correlates well with that found for QDs and proteins of similar size (*26, 27*).

In summary, we have shown that Ni-NTA-coated QDs provide a straightforward method to, in one step, purify and fluorescently reversibly label proteins. By using these QDs, we have selectively purified and labeled an N-terminal His_6-tagged GST, which was not possible using QDs with carboxylates at the surface. Moreover, we have found that Ni^{2+} provides a docking site which helps to precisely orient the fluorescent nanoparticle on the protein surface and that, as a result, GST retained its activity. The use of His-tags has been broadly adopted in the molecular biology and biochemistry communities, and therefore this specific conjugation strategy should enable widespread use of these QDs for a broad range of biological applications.

ACKNOWLEDGMENT

We are grateful to EaStCHEM for a PhD studentship to M.G. A.T.V. acknowledges the Xunta de Galicia (Spain) for a postdoctoral fellowship. A.C. is funded by a Marie Curie Fellowship. We acknowledge support from the EPSRC to purchase the XPS. We would like to thank Dr. Wuzong Zhou and Ross Blackley at EaStCHEM-St. Andrews for the HRTEM studies.

Supporting Information Available: Details on experimental procedures and characterization data. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Medintz, I. L., Uyeda, H. T., Goldman, E. R., and Mattoussi, H. (2005) Quantum dot bioconjugates for imaging, labelling and sensing. *Nat. Mater. 4*, 435–446.
(2) Gao, X., Yang, L., Petros, J. A., Marshall, F. F., Simons, J. W., and Nie, S. (2005) In vivo molecular and cellular imaging with quantum dots. *Curr. Opin. Biotechnol. 16*, 63–72.
(3) Alivisatos, A. P., Gu, W., and Larabell, C. (2005) Quantum dots as cellular probes. *Annu. Rev. Biomed. Eng. 7*, 55–76.
(4) Michalet, X., Pinaud, F. F., Bentolila, L. A., Tsay, J. M., Doose, S., Li, J. J., Sundaresan, G., Wu, A. M., Gambhir, S. S., and Weiss, S. (2005) Quantum dots for live cells, in vivo imaging, and diagnostics. *Science 307*, 538–544.
(5) Klostranec, J. M., and Chan, W. C. (2006) Quantum dots in biological research: recent progress and present challenges. *Adv. Mater. 18*, 1953–1964.
(6) Babu, P., Sinha, S., and Surolia, A. (2007) Sugar-quantum dot conjugates for a selective and sensitive detection of lectins. *Bioconjugate Chem. 18* (1), 146–151.
(7) Howarth, M., Takao, K., Hayashi, Y., and Ting, A. Y. (2005) Targeting quantum dots to surface proteins in living cells with biotin ligase. *Proc. Natl. Acad. Sci. U.S.A. 102*, 7583–7588.
(8) Waugh, D. S. (2005) Making the most of affinity tags. *Trends Biotechnol. 23*, 316–320.
(9) Esposito, D., and Chatterjee, D. K. (2006) Enhancement of soluble protein expression through the use of fusion tags. *Curr. Opin. Biotechnol. 17*, 353–358.
(10) Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol. 60*, 523–533.
(11) Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature 258*, 598–599.
(12) Bornhorst, J. A., and Falke, J. J. (2000) Purification of proteins using polyhistidine affinity tags. *Methods Enzymol. 326*, 245–254.
(13) Guignet, E. G., Hovius, R., and Vogel, H. (2004) Reversible site-selective labeling of membrane proteins in live cells. *Nat. Biotechnol. 22*, 440–444.
(14) Lata, S., Reichel, A., Brock, R., Tampé, R., and Piehler, J. (2005) High-affinity adaptors for switchable recognition of histidine-tagged proteins. *J. Am. Chem. Soc. 127*, 10205–10215.
(15) Abad, J. M., Mertens, S. F. L., Pita, M., Fernandez, V. M., and Schiffrin, D. J. (2005) Functionalization of thioctic acid-capped gold nanoparticles for specific immobilization of histidine-tagged proteins. *J. Am. Chem. Soc. 127*, 5689–5694.
(16) Sigal, G. B., Bamdad, C., Barberis, A., Strominger, J., and Whitesides, G. M. (1996) A self-assembled monolayer for the binding and study of histidine-tagged proteins by surface plasmon resonance. *Anal. Chem. 68*, 490–497.
(17) Cho, M., Lee, S., Han, S.-Y., Park, J.-Y., Rahman, M. A., Shim, Y.-B., and Ban, C. (2006) Electrochemical detection of mismatched DNA using a MutS probe. *Nucleic Acids Res. 34*, e75/1–e75/10.
(18) Hainfeld, J. F., Liu, W., Halsey, C. M. R., Freimuth, P., and Powell, R. D. (1999) Ni-NTA-gold clusters target His-tagged proteins. *J. Struct. Biol. 127*, 185–198.

(19) Kapanidis, A. N., Ebright, Y. W., and Ebright, R. H. (2001) Site-specific incorporation of fluorescent probes into protein: hexahistidine-tag-mediated fluorescent labeling with (Ni^{2+}: nitrilotriacetic Acid)$_n$-fluorochrome conjugates. *J. Am. Chem. Soc. 123*, 12123–12125.

(20) Goldsmith, C. R., Jaworski, J., Sheng, M., and Lippard, S. J. (2006) Selective labeling of extracellular proteins containing polyhistidine sequences by a fluorescein-nitrilotriacetic acid conjugate. *J. Am. Chem. Soc. 128*, 418–419.

(21) Kim, J., Park, H. Y., Kim, J., Ryu, J., Kwon, do. Y., Grailhe, R., and Song, R. (2008) Ni-nitrilotriacetic acid-modified quantum dots as a site-specific labeling agent of histidine-tagged proteins in live cells. *Chem. Commun. 16*, 1910–1912.

(22) Sapsford, K. S., Pons, T., Medintz, I. L., Higashiya, S., Brunel, F. M., Dawson, P. E., and Mattoussi, H. (2007) Kinetics of metal-affinity driven self-assembly between proteins or peptides and CdSe-ZnS Quantum dots. *J. Phys. Chem. C 111*, 11528–11538.

(23) Yao, H., Zhang, Y., Xiao, F., Xia, F., and Rao, J. (2007) Quantum dot/bioluminescence resonance energy transfer based highly sensitive detection of proteases. *Angew. Chem., Int. Ed. 46*, 4346–4349.

(24) Mahajan, S., and Atkins, W. M. (2005) The chemistry and biology of inhibitors and pro-drugs targeted to glutathione S-transferases. *Cell. Mol. Life Sci. 62*, 1221–33.

(25) Shi, B., Stevenson, R., Campopiano, D. J., and Greaney, M. F. (2006) Discovery of glutathione S-transferase inhibitors using dynamic combinatorial chemistry. *J. Am. Chem. Soc. 128*, 8459–8467.

(26) Mattoussi, H., Mauro, J. M., Goldman, E. R., Anderson, G. P., Sundar, V. C., Mikulec, F. V., and Bawendi, M. G. (2000) Self-assembly of CdSe-ZnS quantum dot bioconjugates using an engineered recombinant protein. *J. Am. Chem. Soc. 122*, 12142–12150.

(27) Medintz, I. L., Clapp, A. R., Mattoussi, H., Goldman, E. R., Fisher, B., and Mauro, J. M. (2003) Self-assembled nanoscale biosensors based on quantum dot FRET donors. *Nat. Mater. 2*, 630–638.

(28) Ipe, B. I., and Niemeyer, C. M. (2006) Nanohybrids composed of quantum dots and cytochrome P450 as photocatalysts. *Angew. Chem., Int. Ed. 45*, 504–507.

(29) Lee, I. S., Lee, N., Park, J., Kim, B. H., Yi, Y. W., Kim, T., Kim, T. K., Lee, I. H., Paik, S. R., and Hyeon, T. (2006) Ni/NiO core/shell nanoparticles for selective binding and magnetic separation of histidine-tagged proteins. *J. Am. Chem. Soc. 128*, 10658–10659.

(30) Xu, C., Xu, K., Gu, H., Zhong, X., Guo, Z., Zheng, R., Zhang, X., and Xu, B. (2004) Nitrilotriacetic acid-modified magnetic nanoparticles as a general agent to bind histidine-tagged proteins. *J. Am. Chem. Soc. 126*, 3392–3393.

(31) Lee, K. B., Park, S., and Mirkin, C. A. (2004) Multicomponent magnetic nanorods for biomolecular separations. *Angew. Chem., Int. Ed. 43*, 3048–3050.

(32) Lee, K. S., and Lee, I. S. (2008) Decoration of superparamagnetic iron oxide nanoparticles with Ni^{2+}: agent to bind and separate histidine-tagged proteins. *Chem. Commun. 6*, 709–711.

(33) Kim, S. H., Jeyakumar, M., and Katzenellenbogen, J. A. (2007) Dual-mode fluorophore-doped nickel nitrilotriacetic acid-modified silica nanoparticles combine histidine-tagged protein purification with site-specific fluorophore labeling. *J. Am. Chem. Soc. 129*, 13254–13264.

(34) Habig, W. H., Pabst, M. J., and Jakoby, W. B. (1974) Glutathione S-transferases. The first enzymatic step in mercapturic acid formation. *J. Biol. Chem. 249*, 7130–7139.

(35) McTigue, M. A., Williams, D. R., and Tainer, J. A. (1995) Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. *J. Mol. Biol. 246*, 21–27.

BC800273J